Sciences et Ingénierie de l'Environnement

Chaire de SIRS
Système d'Informations à Référence Spatiale



Travail pratique de diplôme

PROCÉDURES D'AGRÉGATION MULTIFACTORIELLE D'UNITÉS TERRITORIALES

Application à l'expression cartographique de diverses statistiques agricoles



Statistique suisse
7 Agriculture et sylviculture



Lausanne, mars 2002

Candidat	Bertrand BARBEY
Professeur	François GOLAY
Encadrement	Régis CALOZ, EPFL-SIRS
	Jean-François FRACHEBOUD, OFS

TABLE DES MATIÈRES

INDEX DES TABLEAUX 4

INDEX DES FIGURES....5

INDEX DES ÉQUATIONS 5

RÉSUMÉ....6

REMERCIEMENTS....6

LISTE DES ABRÉVIATIONS....7

1 INTRODUCTION....8
- 1.1 Contexte 8
- 1.2 La problématique 8
- 1.3 Objectifs....9
- 1.4 Méthodologie....9

2 THÉORIE.... 11
- 2.1 Systèmes d'Information à Référence Spatiale (SIRS).... 11
  - 2.1.1 DONNÉES 11
  - 2.1.2 MANIPULATIONS....11
  - 2.1.3 ANALYSE SPATIALE....11
- 2.2 ArcView – Avenue.... 12
- 2.3 Méthodes de relevés statistiques 12
- 2.4 Etablissement d'une typologie des entreprises agricoles 13
- 2.5 Principes de la résolution multicritère 15
- 2.6 Présentation et évaluation de quelques méthodes statistiques 15
  - 2.6.1 TESTS SUR LES OBSERVATIONS 16
  - 2.6.2 TESTS SUR LES RANGS....17
  - 2.6.3 ANALYSE EN COMPOSANTES PRINCIPALES ACP....17

3 INFORMATIONS PRÉLIMINAIRES 19
3.1 Office Fédéral de la Statistique 19
3.1.1 FONCTIONNEMENT GÉNÉRAL 19
3.1.2 DANS LE CADRE DE CE PROJET 19
3.2 Données à disposition 19
3.3 Format des données 20
3.4 Besoins 21
4 CRITÈRES D'AGRÉGATION 22
4.1 Critère de confidentialité 23
4.2 Critère de ressemblance 23
4.3 Critère d'appartenance au canton 24
4.4 Critère d'esthétisme 24
4.4.1 L'INDICE DE COMPACITÉ 24
4.4.2 LA FRONTIÈRE COMMUNE 25
4.4.3 LA TAILLE DE L'AGRÉGAT 26
5 PRINCIPALES ÉTAPES D'AGRÉGATION 27
5.1 Description théorique 27
5.2 Simulation d'agrégation 28
6 SITUATIONS CONTENTIEUSES 30
6.1 Mauvaise agrégation du point de vue de la forme de l'agrégat 30
6.2 Mauvaise agrégation du point de vue de la ressemblance 31
6.3 Le cas particulier du Tessin 33
7 SIGNIFICATION DES INDICATEURS 37
7.1 Appartenance au canton 37
7.2 Ressemblance 37
7.3 Esthétisme 38
8 AUTRES PROCÉDURES D'AGRÉGATION 41
8.1 Méthodes statistiques 41
8.1.1 CORRÉLATION SIMPLE SUR LES OBSERVATIONS 41
8.1.2 CORRÉLATION SUR LES RANGS 42
8.2 Calcul d'une distance 42
8.2.1 SCORE DE PEARSON (KHI-CARRÉ) 42
8.2.2 DISTANCE SUR LES COMPOSANTES PRINCIPALES 43

9 DÉVELOPPEMENT DU PROTOTYPE SUR LES DONNÉES 2000 44
9.1 Méthode Multicritère 44
9.1.1 TYPOLOGIE 44
9.1.2 SPÉCIALISATION 45
9.2 Méthode de la corrélation sur les rangs 47
9.3 Méthode par calcul de distance 48
9.3.1 DISTANCE SUR LES MBS EN POURCENT 49
9.3.2 DISTANCE SUR LES COMPOSANTES PRINCIPALES 49
10 GÉNÉRALISATION 51
10.1 Combinaison de méthodes 51
10.2 Application de l'agrégation de communes sur d'autres données 54
11 SYNTHÈSE 55
11.1 Les avantages d'un programme d'agrégation d'unités territoriales 55
11.2 Inconvénients et limitations 55
11.3 Tableau synthétique des méthodes testées 57
12 PERSPECTIVES 58
12.1 Optimisation du prototype 58
12.2 Marchés potentiels 58
12.3 Modifications de la classification 59
12.4 Affranchissement des limites communales 59
12.5 Autre transformation des données du recensement 60
13 CONCLUSION 61
14 BIBLIOGRAPHIE 62

INDEX DES TABLEAUX

Tableau 1 : Définition des orientations et classes de production (spécialisation)....................13
Tableau 2 : Méthode multicritère : calcul des scores...15
Tableau 3 : Récapitulatif des données à disposition (description et format) 20
Tableau 4 : Présentation des critères et indicateurs de la méthode multicritère appliquée aux données de 1996 22
Tableau 5 : Résultats comparatifs de deux méthodes sur l'ensemble du Tessin 36
Tableau 6 : Critères et indicateurs retenus pour les applications sur les données 2000...........40
Tableau 7 : Corrélation simple : résultats d'une itération, avec seuil de corrélation variable ...41
Tableau 8 : Corrélation simple : résultats comparatifs avec la méthode multicritère 41
Tableau 9 : Corrélation sur les rangs : résultats d'une itération: seuil de corrélation variable..42
Tableau 10 : Corrélation sur les rangs : résultats comparatifs avec la méthode multicritère....42
Tableau 11 : Scores de Pearson : résultats d'une itération..44
Tableau 12 : ACP : résultats pour les variables servant au calcul de la typologie...................43
Tableau 13 : Données 2000 : Récapitulatif des critères et indicateurs....................................44
Tableau 14 : Méthode multicritère sur la typologie : variations des poids et détermination du jeu de référence 45
Tableau 15 : Méthode multicritère sur la spécialisation...46
Tableau 16 : Corrélation de Spearman : variations du seuil de corrélation : exemple 1..........47
Tableau 17 : Corrélation de Spearman : variations du seuil de corrélation : exemple 2..........48
Tableau 18 : Calcul de distance euclidienne selon différents schémas, fonctions de la typologie ou de la spécialisation du germe 49
Tableau 19 : ACP : résultats pour les marges brutes standard primaires sur les exploitations .50
Tableau 20 : Combinaison de méthodes : multicritère + distance ou corrélation....................52
Tableau 21 : Comparaison des méthodes selon le nombre moyen de communes et de types de production différents regroupés dans chaque agrégat 53
Tableau 22 : Nombre d'agrégats, indice de Gravélius et nombre d'exploitations moyens par agrégat : sur toute la Suisse et sur les régions où les différentes méthodes n'ont pas effectué les mêmes agrégations...53
Tableau 23 : Catégories pour la représentation des surfaces herbagères 54
Tableau 24 : résultats des deux méthodes retenues pour l'agrégation de communes pour la cartographie des surfaces herbagères en rapport à la SAU totale. 54
Tableau 25 : Tableau comparatif des méthodes testées 57

INDEX DES FIGURES

Figure 1 : Schéma de relevés statistiques 12
Figure 2 : Application de l'indice de Gravélius à l'hydrologie 25
Figure 4 : Agrégat de forme irrégulière : exemple 1 30
Figure 5 : Agrégat de forme irrégulière : exemple 2 31
Figure 6 : Situation litigieuse du point de vue de la ressemblance 31
Figure 7 : Influence du critère d'appartenance au canton 32
Figure 8 : Commune isolée du point de vue de la typologie : exemple 1 32
Figure 9 : Commune isolée du point de vue de la typologie : exemple 2 33
Figure 10 : Tessin, cas particulier sur les bords du lac de Lugano 33
Figure 11 : Tessin : variations de la VL sur le nombre d'exploitations, ...34
Figure 12 : Tessin : alternatives avec pondération uniforme 35
Figure 13 : Tessin : alternatives avec un poids fort sur l'indicateur de taille 35
Figure 14 : Tessin : variantes avec indicateurs d'esthétisme prédominants 36
Figure 15 : Conflits entre agrégation basée sur la spécialisation 46

INDEX DES ÉQUATIONS

Équation 1 : calcul du score de Pearson 16
Équation 2 : calcul du coefficient de corrélation sur les observation 16
Équation 3 : Calcul du coefficient de Spearman : corrélation sur les rangs 17
Équation 4 : indice de compacité de Gibbs 24
Équation 5 : indice de compacité de Cole 24
Équation 6 : indice de compacité de Gravélius 24

RÉSUMÉ

Ce travail de diplôme associe l'Office fédéral de la statistique au laboratoire de SIRS de l'EFPL. L'OFS désire trouver une solution pour la diffusion cartographique de ces statistiques agricoles respectant les contraintes fixées par la loi sur la protection des données. Un prototype est développé à partir des données d'orientation de production des exploitations agricoles, le but étant de modifier le moins possible la typologie des communes réunies.

Ce projet développe, à l'aide des technologies des SIT, un programme permettant l'agrégation de polygones, représentant les communes suisses, en suivant une première condition d'adjacence. Pour les solutions retenues, le deuxième facteur d'agrégation consiste en une comparaison multicritère des communes voisines d'un noyau sélectionné, les critères étant définis à partir des attributs ou de la géométrie des polygones.
Parmi les deux systèmes aboutissants aux résultats les plus concluants, la procédure intégrant un calcul de distance pour le critère de ressemblance et une résolution multicritère globale ouvre de nouvelles perspectives et élargit son champ d'application à d'autres types de données statistiques.

D'autre part, ce travail expose différentes approches pour le choix des agrégations à effectuer, les limites des solutions retenues, leurs avantages et inconvénients, ainsi que les possibilités d'amélioration et d'utilisation de telles procédures.

REMERCIEMENTS

J'aimerais faire part de ma gratitude à toutes les personnes qui, de près ou de loin, m'ont permis de mener à bien ce travail de diplôme.

Mes remerciements s'adressent particulièrement aux personnes suivantes :

M. François Golay, professeur responsable de la chaire de SIRS qui m'a permis de réaliser ce travail intéressant et gratifiant.

M. Régis Caloz, collaborateur et chargé de cours à la chaire de SIRS, pour son encadrement, sa disponibilité et ses conseils constructifs.

M. Jean-François Fracheboud, directeur de la Section d'agriculture et de sylviculture de l'OFS, pour sa participation active à l'élaboration du sujet. Ses informations et son attention quant à l'évolution du projet ont été une source de motivation constante.

Les collaborateurs de l'OFS, Nadia Camilli, Nadia Rognon, Daniel Bohnenblust et Hans Steffen, pour leur accueil, leur disponibilité et la qualité de leurs interventions ayant permis une progression continue du projet.

Les collaborateurs de la chaire de SIRS, Daniel Gnerre pour son soutien dans l'apprentissage du langage de programmation Avenue, et Marc Gilgen pour son intérêt et sa disponibilité.

LISTE DES ABRÉVIATIONS

AGIR	Agence d'Information Agricole Romande
CAO	Conception Assistée par Ordinateur
CEE	Communauté Economique Européenne
Cst	Constitution Suisse
EMG	Espace Média Groupe
EPFL, SSIE	Ecole Polytechnique Fédérale de Lausanne, Section Sciences et Ingénierie de l'Environnement
ESRI	Environmental Systems Research Incorporation
FAO	Food and Agriculture Organization of the United Nations
FAT	Station Fédérale de recherche en économie et technologies agricoles de Tänikon
GEOSTAT	SIT géré par l'OFS et contenant leurs données statistiques
LAgr	Loi Fédérale sur l'Agriculture
LFC	Longueur de la Frontière Commune entre un germe et sa commune voisine
LPD	Loi Fédérale sur la Protection des Données
LSF	Loi sur la Statistique Fédérale
MBS	Marge Brute Standard, gain potentiel normalisé pour chacune des diverses productions agricoles
MN03	Campagne de Mensuration Nationale entreprise sur la base de l'ellipsoïde de Bessel reflétant le système universel WGS en 1903
MNA	Modèle Numérique d'Altitudes
MS	Mobilité Spatiale
OCDE	Organisation de Coopération et de Développement Economiques
ODA	Ordonnance sur le relevé et le traitement des Données Agricoles
OFAG	Office Fédéral de l'Agriculture
OFEFP	Office Fédéral pour la protection de l'Environnement, de la Forêt et du Paysage
OFS / BFS	Office Fédéral de la Statistique / Bundesamt Für Statistik
OLPD	Ordonnance d'application de la LPD
PSL	Producteurs Suisses de Lait
SA GmbH	Schweizer Agrarmedien GmbH
SAR	Swiss Agricultural Research, réseau de 6 stations de recherche agricole
SAU	Surface Agricole Utile
SIRS	Système d'Informations à Référence Spatiale
SIG / SIT	Système d'Informations Géographiques / du Territoire
UE	Union Européenne
USP	Union Suisse des Paysans
VL	Valeur Limite du nombre d'exploitations pour vérifier la contrainte de confidentialité

1 INTRODUCTION

1.1 CONTEXTE

L'agriculture connaît aujourd'hui des mutations considérables. Les enjeux des réorientations en cours des activités agricoles sont importants : maintien d'une agriculture concurrentielle, aspects paysagers, peuplement décentralisé du territoire, etc. Par ailleurs, l'évolution est notable à tous les niveaux de la société actuelle. L'Office Fédéral de la Statistique (OFS) dans son intégralité doit alors tenir compte de nouveaux paramètres contraignants lors de chaque étape du relevé, du traitement et de la diffusion de l'information. Par exemple, l'OFS est soumis à la loi (1992) et à l'ordonnance (1993) sur la protection des données selon lesquelles "les résultats du traitement sont publiés sous une forme ne permettant pas d'identifier les personnes concernées" (LPD, art. 22). De même, les négociations bilatérales entre l'Union Européenne et la Suisse en matière d'agriculture induisent des ajustements dans l'approche des recensements des structures agricoles par exemple.

Les données statistiques alors relevées sur les exploitations agricoles reflètent un état, une situation figée de l'économie, du fonctionnement et de la structure des entreprises du secteur primaire. En rassemblant les informations de plusieurs recensements, on dispose de données de dimensions spatiale et temporelle permettant toutes sortes d'analyses, pour juger de l'efficacité des actions entreprises, pour comprendre des processus de changements naturels ou pour décider des mesures à prendre dans le cadre de nouveaux projets. Toutes ces réflexions doivent permettre à l'agriculture de remplir au mieux ses différentes fonctions et de maintenir son importance dans l'économie contemporaine.

Les informations saisies lors des enquêtes, ainsi que les résultats de toutes les manipulations de données sont souvent exposés dans des tableaux et des graphiques. Une grande part des statistiques agricoles ayant un lien direct avec l'espace, la section d'Agriculture et Sylviculture de l'OFS (OFS Agr) présente de nombreuses données sous forme de cartes. Une part de la transformation des relevés en information cartographique s'effectue encore manuellement, mais l'évolution technologique permet d'ouvrir de nouveaux horizons dans ce domaine. Au niveau des relevés statistiques, l'avènement du numérique accroît la rapidité et l'efficacité des opérations d'enregistrement, d'analyse et de transfert de données. Les Systèmes d'Informations du Territoire (SIT) proposent une démarche et des outils adaptés pour la gestion et la présentation des informations revêtant un caractère spatial.

1.2 LA PROBLEMATIQUE

L'OFS enregistre de nombreux paramètres sur les exploitations du secteur primaire, et il en combine parfois pour la diffusion en thèmes porteurs tels que la main d'œuvre, l'orientation de production ou le bétail bovin. En outre, la confidentialité des informations collectées par l'OFS doit être respectée lors de leur diffusion. La plus petite entité politique et administrative, la commune, s'impose souvent pour la présentation des résultats sous forme de graphiques ou de cartes. Pourtant, cette entité spatiale n'est pas toujours appropriée pour les statistiques agricoles, pas plus d'ailleurs que d'autres découpages existants comme le district par exemple. Il arrive fréquemment qu'une commune ne réunisse pas suffisamment d'exploitations agricoles pour garantir l'anonymat des agriculteurs comme dans des cas extrêmes où une seule exploitation se trouve sur le territoire communal. Pour garantir alors

les objectifs de la protection des données, la solution actuellement mise en œuvre consiste à regrouper des communes en un agrégat et à présenter une statistique englobant les données de toutes les communes ainsi réunies. La plupart des opérations cartographiques transformant les relevés en données anonymes s'effectuent encore manuellement, à l'heure où le développement des SIT permet d'envisager une assistance informatique efficace. L'OFS Agr s'intéresse d'ailleurs fortement à ces nouvelles techniques. La chaire de SIRS du Département de Génie Rural (DGR) adhère à ce projet en espérant qu'une telle méthode puisse également servir à d'autres domaines de la statistique ou de la cartographie.

1.3 OBJECTIFS

Pour garantir des données statistiques fiables et représentatives, l'OFS Agr doit s'adapter aux changements induits, entre autres, par la nouvelle politique agricole (PA2002). Des réorganisations sur les systèmes d'acquisition et de traitement des données sont d'ailleurs prévues dans le programme pluriannuel (1999-2003) de la statistique fédérale suite à la révision des textes législatifs dans le domaine agricole. Nous pouvons mentionner l'Ordonnance sur le relevé et le traitement de Données Agricoles (ODA, 1998) et les textes législatifs pris en considération dans le cadre des négociations bilatérales avec l'Union Européenne comme principales sources de modifications.

D'autre part, l'OFS Agr désire également utiliser le potentiel des nouvelles technologies dans le domaine des SIRS afin d'obtenir une assistance informatique à l'élaboration de ses cartes. L'objectif fondamental de ce travail est de proposer une ou plusieurs solutions pour automatiser autant que possible la phase d'agrégation de communes dans l'optique de diffusion cartographique des informations de la statistique agricole. Deux contraintes principales mettent une orientation et un cadre plus précis à ce travail : il s'agit tout d'abord de préserver la confidentialité des données en respectant un nombre minimal d'exploitations dans les communes, et ensuite d'assurer un aspect visuel clair en s'occupant de l'esthétisme de la carte (forme et taille des agrégats).

Dans un esprit plus global, nous désirons évoquer ce que peuvent apporter les moyens actuels fournis par les SIT à la problématique générale de l'OFS et donner quelques pistes quant à la réalisation de ces perspectives. Par exemple, il est légitime de vouloir utiliser les fonctionnalités des SIT pour déterminer de nouveaux indicateurs en combinant les informations de l'OFS avec des données de domaines variés, comme la topographie, la pédologie ou la météorologie. Il s'agit également d'ouvrir la discussion sur les possibilités de s'affranchir des frontières administratives pour la représentation des statistiques agricoles.

1.4 METHODOLOGIE

Ce travail est essentiellement organisé sur la résolution de l'agrégation des communes ne vérifiant pas la contrainte de confidentialité. Il se déroule en quatre phases principales : la poursuite du travail de R.Tornay (2001) sur les données du recensement 96, l'adaptation et le développement du prototype sur les données 2000 d'orientation de production, la mise en place d'un système plus général et une étape d'analyses et de perspectives.
Pour la première phase, il s'agit tout d'abord de déterminer un système d'indicateurs et de paramètres correspondants, en tenant compte des contraintes informatiques et des vœux exprimés par l'OFS Agr.
Ceux-ci déterminent trois axes d'agrégation :

1. respect de la protection des données
2. ressemblance entre communes, basée sur la variable à représenter
3. esthétisme du résultat

Ensuite, nous programmons une solution traitant tous les facteurs susmentionnés de manière intégrée, soit une procédure d'agrégation multicritère laissée en suspens lors du travail précédent. En effet, ce type de cheminement s'apparente beaucoup aux procédés manuels mis en place actuellement. Il s'agit alors de déterminer une pondération adéquate des différents paramètres. A ce stade, il s'agit de déceler les cas particuliers qui peuvent entraver le déroulement de l'agrégation et d'y apporter les solutions adéquates. C'est sur de telles configurations de polygones que nous devons orienter nos recherches afin de déterminer un jeu de poids de référence permettant d'obtenir l'agrégation la plus favorable. De plus, dans une première tentative d'amélioration du programme, nous évaluons la pertinence des indicateurs choisis en comparant diverses exécutions de la procédure. Finalement, nous effectuons une première approche des méthodes statistiques et de calcul de distance pouvant se révéler utiles dans ce projet, du point de vue d'une ouverture possible de cette méthode à d'autres domaines de la statistique.

Pour la deuxième phase du projet, nous disposons d'un jeu de données plus complet permettant d'approfondir et d'élargir la problématique, en poursuivant le développement des méthodes sélectionnées dans l'étape précédente, et en testant un système basé sur un calcul de distance, a priori bien adapté à ce genre de situations. Nous maintenons le procédé de résolution multicritère comme référence, explorons les possibilités offertes par une méthode de corrélation sur les rangs et examinons les ressources offertes par la détermination d'une distance sur les variables caractérisant les communes.

Dans l'étape suivante, nous nous concentrons sur les possibilités de généralisation de ce programme, en sachant que de telles considérations interviennent déjà lors du développement du prototype. Il s'agit d'évaluer une méthode consistant à combiner deux des systèmes testés auparavant, pour garder le traitement intégré de plusieurs facteurs tout en laissant totalement libre le choix des variables à cartographier. Nous optons pour une intégration du calcul de distance ou de la corrélation dans le processus de résolution multicritère.

Finalement, en quatrième partie, nous effectuons une synthèse des méthodes testées et développons les avantages et les inconvénients liés aux solutions les plus performantes. D'autre part, si le temps le permet, nous nous attacherons à relever les moyens dont disposent actuellement les SIT et qui peuvent permettre à ce prototype d'évoluer vers un outil intégré de traitement et de cartographie de données statistiques. Nous pourrons envisager quelques perspectives pour le traitement des données de l'OFS, mais également quelques idées sur leur combinaison avec d'autres couches d'informations.

2 THEORIE

2.1 SYSTÈMES D'INFORMATIONS À RÉFÉRENCE SPATIALE

2.1.1 DONNEES

Le but d'un SIRS est de présenter le monde réel sous la forme d'un modèle. Pour atteindre cet objectif, deux types de données sont nécessaires :

- la géométrie :
  - Dans le modèle vectoriel, les éléments de base sont concentrés en trois catégories : entités ponctuelle, linéaire ou surfacique. Chaque objet est alors caractérisé par un ensemble de coordonnées géographiques représentant les points fondamentaux (sommets d'un polygone, début et fin d'une ligne…)
  - Le modèle raster est constitué par une grille régulière dont chaque élément (pixel) contient une part de l'information sur un thème sous la forme d'un code. Chaque pixel possède une géoréférence, soit les coordonnées d'un point caractéristique et la taille de la maille.
- la thématique :

  La description des objets du modèle et les informations non géométriques sont regroupées dans un tableau annexe, un extrait de base de données. Les tables sont reliées aux entités graphiques qu'elles référencent par l'information géographique, et entre elles par des champs communs.

2.1.2 MANIPULATIONS

Les SIRS intègrent de nombreux outils permettant de manipuler toutes les données pour les rendre cohérentes et ne garder que celles qui sont essentielles au projet. Citons à titre d'exemple la possibilité de combiner visuellement – et physiquement dans une table de synthèse – plusieurs couches d'informations ayant en commun la référence spatiale. En outre, pour de nombreuses opérations géographiques, la finalité consiste à bien visualiser des cartes et des graphes. La carte est en effet un formidable outil de synthèse et de présentation de l’information. Les SIRS offrent à la cartographie moderne de nouveaux modes d’expression permettant d’accroître de façon significative son rôle informatif. Les cartes créées avec un SIRS peuvent désormais facilement intégrer des rapports, des vues 3D, des images photographiques et toutes sortes d’éléments multimédia.
Par la transformation et le traitement des données, il est possible d'établir des liens logiques entre des changements observés et les mesures entreprises pour enrayer ou favoriser un phénomène, naturel ou autre. Ce genre d'examen peut servir de base à des décisions pour des domaines aussi variés que l'administration, l'économie ou la recherche.

2.1.3 ANALYSE SPATIALE

L’intégration de données au travers des différentes couches d’informations permet d’effectuer une analyse spatiale rigoureuse. Cette analyse par croisement d’informations, si elle peut s’effectuer visuellement (à l’identique de calques superposés) nécessite souvent la liaison avec des données alphanumériques. Croiser la nature d’un sol, sa déclivité, la végétation présente avec les propriétaires et les subventions allouées constitue un exemple d’analyse sophistiquée que permet l’usage d’un SIRS. Nous tendons à développer une facette de cette fonctionnalité dans le cadre de

ce travail puisqu'il s'agit de comparer les polygones adjacents sur la base de leurs attributs respectifs.

2.2 ARCVIEW – AVENUE

ArcView est l'un des logiciels de SIRS les plus utilisés et les plus puissants pour la résolution de problèmes touchant à l'analyse spatiale et à la cartographie. Il permet d'accéder aux données de fichiers aux formats les plus divers : géométrie et attributs de ArcView, ArcInfo et MapInfo, images (TIFF, JPEG, BMP…), bases de données (EXCEL…), dessins CAO (AutoCAD, MicroStation…), etc. De cette technologie, nous utilisons la version ArcView 3.2 pour des raisons de compatibilité entre l'EPFL et l'OFS.
Le langage Avenue est l'une des méthodes les plus accessibles et les plus utilisées pour l'élaboration de programmes et applications à exécuter avec le logiciel ArcView. Tout comme Visual Basic, Avenue est un langage de programmation orienté objet très simple de structure et d'utilisation, compatible avec le logiciel ArcView. Comme le premier prototype a été réalisé avec ce langage, il est apparu évident de poursuivre dans cette voie, bien que la nouvelle version, ArcGIS 8.1, utilise préférentiellement le langage Visual Basic.

2.3 MÉTHODES DE RELEVÉS STATISTIQUES

L'OFS suit les principes suivants pour l'établissement des informations statistiques :

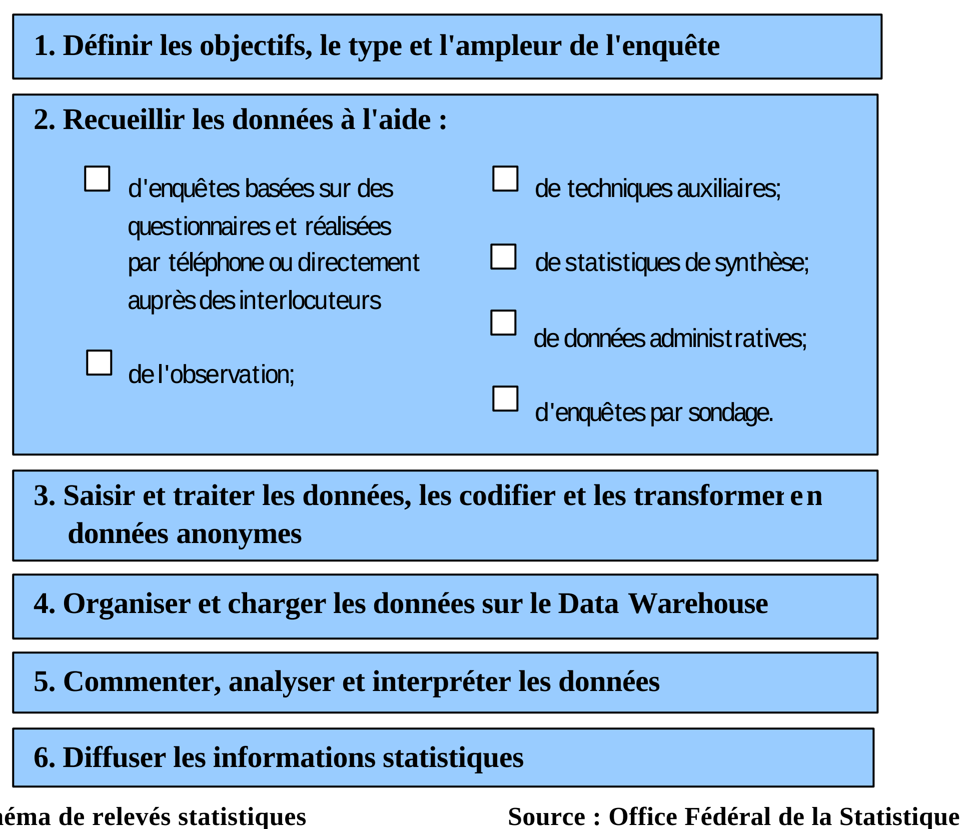


Figure 1 : Schéma de relevés statistiques **Source : Office Fédéral de la Statistique**

La statistique des structures agricoles se fonde sur un catalogue de variables établi par l'OFS et conforme aux exigences européennes (Eurostat). Les cantons sont chargés de la collecte des données et utilisent à cet effet différents questionnaires, fédéraux ou cantonaux.

Ensuite, les cantons traduisent les données récoltées sous forme numérique selon un schéma prédéfini et les transmettent à l'OFS qui effectue plusieurs traitements sur ces informations :

- vérification de l'exhaustivité des données
- vérification de l'exactitude des données déclarées par l'exploitant et corrections nécessaires.
- calcul des marges brutes par exploitation et par type de production
- détermination de la spécification de chaque entreprise
- regroupement de ces données par exploitation dans la base de données OFS : 1 ligne par exploitation, avec toutes les caractéristiques, mais sans l'identification des agriculteurs
- agrégation des résultats par communes, districts ou cantons, selon les commandes

Remarque sur la qualité des informations :

Les marges brutes standard (MBS) par type de production sont calculées par la FAT, la Station fédérale de recherche en économie et technologie agricole de Tänikon, qui fait partie d'un groupe de 6 stations de la recherche agricole suisse (SAR, Swiss Agricultural Research). La FAT procède à une mise en valeur centralisée de données comptables issues d'un échantillon de 3000 à 4000 exploitations et calcule, sur cette base, une série de chiffres clés qui fourniront une information importante dans le cadre de l'évaluation de l'impact des différentes mesures de politique agricole, sachant que les paiements directs occupent le premier plan dans le revenu de l'agriculteur. De plus, les résultats sont également mis à disposition pour la recherche, la formation, la vulgarisation, l'estimation des biens-fonds agricoles, la prise de décisions agro-politiques y compris l'évaluation des mesures de politique agricole. Pour l'OFS, la FAT établit des normes ou standards sur chaque type de production agricole, comme par exemple, la marge brute standard (MBS) pour 1 hectare d'orge. Cette valeur correspond au rendement financier potentiel de la culture en question dont sont soustraites les dépenses directement liées (semences, engrais…). La FAT donne également des valeurs théoriques sur la main d'œuvre nécessaire au fonctionnement d'une exploitation agricole, selon l'orientation de production et en considérant uniquement le travail de l'exploitant et de sa famille. Cette valeur contribue à distinguer les entreprises exploitées à titre principal ou accessoire.

Les données de 1996 transmises par l'OFS sont le fruit d'une étape d'agrégation supplémentaire. Toutes les données du recensement des structures agricoles traduites en terme monétaire (MBS) sont regroupées dans les codes D01 à J18 dont le détail est présenté en annexe I.1. Ensuite, une première simplification synthétise ces données en 5 catégories principales P1 à P5 et 7 secondaires P11 à P131 (cf. annexe I.1). A partir des 5 thèmes primordiaux est déterminée la spécialisation de chaque exploitation en 3 orientations et 8 classes :

Orientation	Classes
Production végétale	1. Grandes cultures 2. Cultures horticoles 3. Cultures permanentes
Production animale	4. Herbivores 5. Granivores (Elevage hors sol)
Exploitation mixte	6. Polyculture 7. Polyélevage 8. Mixte

Tableau 1 : Définition des orientations et classes de production (spécialisation)

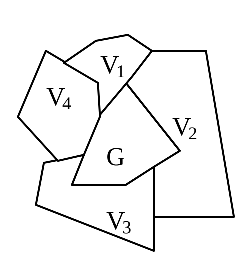
2.4 ETABLISSEMENT D'UNE TYPOLOGIE DES ENTREPRISES AGRICOLES

Finalement, l'information sur toutes les exploitations est rendue anonyme en définissant une typologie générale sur la commune. Cette typologie est un code (15 catégories, codes 1 à 99), exprimant d'une part quelle orientation est représentée en majorité sur la commune, et d'autre part sa proportion par rapport au nombre total d'entreprises agricoles. Le calcul se fait sur la base du nombre d'exploitations spécialisées dans chacune des classes du tableau 1. On classe les entreprises agricoles dans l'une des quatre catégories suivantes : les exploitations spécialisées dans les cultures permanentes, les autres cultures végétales, les productions animales et les exploitations mixtes. Ensuite, on compte combien d'exploitations regroupe chaque catégorie, et on retient celle qui en contient le plus. Enfin, le code de la typologie est décidé selon le nombre d'exploitations (en % du total communal) appartenant à la catégorie prédominante susmentionnée. La description détaillée des typologies et spécialisations est présentée en annexe II.1.

Cette classification des entreprises a été introduite selon le modèle européen qui se fonde sur la contribution des diverses branches de production à la formation de la marge brute standard totale de l'exploitation. Cette méthode permet de comparer l'orientation technico-économique des entreprises sur le plan régional, national et international. La prise en considération d'indicateurs économiques rend possible une typologie évolutive et dynamique, mais présente l'inconvénient d'être dépendante des fluctuations de prix. En effet, la spécialisation des exploitations est calculée sur la base des MBS primaires regroupant toutes les marges brutes estimées pour chaque type de production. Celles-ci sont déterminées comme le rendement financier de la production en question moins les frais directement liés. Déjà au niveau du chiffre d'affaire, les variations peuvent se révéler importantes selon les prix d'achat et de vente. Ensuite, les frais d'exploitation, dont sont exclus le travail de l'exploitant et les amortissements sur les bâtiments et les machines, peuvent subir de fortes fluctuations. Illustrons ces propos par un exemple factice exagéré d'une exploitation menant parallèlement l'engraissage porcin et la culture céréalière. En supposant que le prix du kilogramme de porc à l'exploitant baisse d'environ 50 centimes d'une année à l'autre. A raison de 100kg par cochon, et 600 animaux engraissés durant l'année, la différence de revenu annuel s'élève déjà à Fr. 30'000.-, en considérant stable le prix des porcelets à l'arrivée dans l'exploitation. Dans l'hypothèse où cet exploitant utilise son domaine pour des cultures qui, dans la même période, suivent une évolution de prix inverse, la structure du revenu risque de changer de proportion. Si, de surcroît, le prix de l'aliment pour les animaux monte et que les frais d'achat de semences et d'engrais pour les champs diminuent, il se peut que l'exploitation change de spécialisation, la marge liée aux animaux devenant alors inférieure à celle produite par les céréales. Cette situation traduit deux tendances contradictoires. D'un côté, l'évolution de la structure du revenu agricole de l'exploitation est parfaitement transmise. Par contre, l'ennui réside dans le fait que, d'une année à l'autre, l'exploitation n'a enregistré aucun changement de quantité dans sa production, ce que laisserait supposer la modification de spécialisation.

2.5 PRINCIPES DE LA RÉSOLUTION MULTICRITÈRE

La méthode multicritère représente l'exemple type de la résolution multifactorielle d'un problème. Elle permet d'évaluer simultanément plusieurs variantes en les comparant sur divers paramètres avec la possibilité de conférer des priorités ou importances différentes aux critères retenus en leur attribuant un facteur multiplicatif (poids). Critère après critère, nous évaluons le degré de ressemblance en attribuant un score choisi dans une échelle fixée auparavant par l'opérateur. Dans la situation qui nous occupe, il s'agit de trouver, pour une commune comptant moins de 12 exploitations agricoles (germe), la commune contiguë la plus appropriée pour l'agrégation, à savoir celle qui présente le plus de similitudes sur la structure de production et la forme de l'agrégat la plus régulière.



Dans le petit exemple ci-contre, nous considérons G comme la commune germe et V_i les communes agrégeantes potentielles. Imaginons deux critères C_1 et C_2 dont les poids respectifs sont $P_1 = 1$ et $P_2 = 3$. Sur le premier critère, le score peut prendre les valeurs 2 ou 0, sur le second, les valeurs 2, 1 et 0, selon que le germe et le voisin en question sont très ou pas semblables.

On obtient alors le tableau récapitulatif suivant (factice)

Voisins	**V_1**	**V_2**	**V_3**	**V_4**
Critère C_1	Score $S_{11} = 2$	$S_{21} = 0$	$S_{31} = 0$	$S_{41} = 2$
Critère C_2	$S_{12} = 0$	$S_{22} = 2$	$S_{32} = 0$	$S_{42} = 1$
T_{i1} = Score S_{i1} * P_1	2 * 1 = 2	0 * 1 = 0	0 * 1 = 0	2 * 1 = 2
T_{i2} = Score S_{i2} * P_2	0 * 3 = 0	2 * 3 = 6	0 * 3 = 0	1 * 3 = 3
Score total = $T_{i1} + T_{i2}$	2 + 0 = 2	0 + 6 = 6	0 + 0 = 0	2 + 3 = 5

Tableau 2 : Méthode multicritère : calcul des scores

Le 2[e] voisin a le score le plus élevé, il va être retenu pour l'agrégation

2.6 PRÉSENTATION ET ÉVALUATION DE QUELQUES MÉTHODES STATISTIQUES

Dans le cadre de ce projet, nous sommes amenés à comparer des polygones, représentant les communes, sur la base de leurs attributs. En particulier, nous devons sélectionner, parmi les communes adjacentes, celle qui ressemble le plus au germe, en regard de quelques attributs caractéristiques regroupés dans un vecteur. Nous considérons alors chacun de ces vecteurs comme un échantillon représentatif d'une distribution statistique, et décidons de tester s'ils proviennent de la même distribution. Hubert Béguin (1979) s'attache principalement à la description démographique, mais il propose quelques tests d'ajustement permettant d'évaluer la signification d'un échantillon d'informations distribuées spatialement. D'autre part, des livres traitant de statistiques offrent des alternatives fréquemment utilisées. (T. & R. Wonnacott, 1991; S. Morgenthaler, 1997)

2.6.1 TESTS SUR LES OBSERVATIONS

2.6.1.1 Kolmogorov – Smirnov

Ce test compare deux échantillons ordonnés sur la base d'un score reflétant la plus grande différence entre les deux distributions adaptées des échantillons. Cependant, il n'est pas recommandé lorsqu'un échantillon contient plusieurs valeurs identiques, ce qui est très fréquent avec les données sur les communes où les entreprises agricoles se regroupent en 3 ou 4 spécialisations sur les 8 catégories définies, d'où la présence de plusieurs valeurs nulles.

2.6.1.2 Khi-carré (Pearson)

Ce test établit une comparaison entre 1 échantillon observé et 1 échantillon de référence (à n variables), ce dernier ayant une distribution théorique attendue ou connue. Il est possible d'adapter ce test à notre situation, en posant comme hypothèse que les données du germe constituent l'échantillon de référence. Un score χ_0^2 proportionnel à la somme des carrés des différences doit être comparé aux tables de la loi khi-carré pour (n-1) degrés de liberté pour vérifier si les distributions sont significativement différentes, au seuil de confiance (1-α), par exemple 90%.

$$c_0^{\;2} = \sum_{i=1}^{n} \left(\frac{(V_i - G_i)^2}{G_i} \right) \qquad V_i \text{ et } G_i : \textit{valeurs du voisin, respectivement du germe}$$

Équation 1 : calcul du score de Pearson

Sous cette forme, ce test n'est pas très utile pour notre problématique : nous voulons déterminer le voisin le plus ressemblant, et non écarter les moins semblables. En effet, pour ce genre de test, on pose une hypothèse H_0 (dite hypothèse nulle) qu'il s'agit de vérifier ou d'infirmer. En comparant le score aux valeurs des tables, on peut tirer l'une des deux conclusions suivantes. On rejette H_0, ce qui signifie que l'on est sûr à plus de 90% que les distributions ne sont pas semblables, ou on ne peut pas rejeter H_0, ce qui revient à dire qu'il y a de bonnes chances pour que les échantillons se ressemblent, mais sans pouvoir préciser quel degré de similitude les réunit.
Par contre, nous pouvons utiliser ce score χ_0^2 dans le cadre de notre projet comme une mesure de distance entre le germe et un voisin.

2.6.1.3 Corrélation simple

Comme pour les deux méthodes précédentes, la corrélation calcule la ressemblance entre deux échantillons et l'exprime dans une échelle allant de -1 à 1. Plus le score s'approche des bornes, plus les comportements se ressemblent, plus il est proche de 0, plus les échantillons sont indépendants. On détermine le coefficient de corrélation à l'aide de la formule :

$$r = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2}\,\sqrt{(Y_i - \overline{Y})^2}} \qquad \textit{avec } \overline{X} \textit{ et } \overline{Y}, \textit{ les moyennes des échantillons respectifs}$$

Équation 2 : calcul du coefficient de corrélation sur les observation

Cette méthode peut se révéler utile, puisqu'elle quantifie le degré de similitude entre les échantillons, ce qui permet de hiérarchiser les voisins adjacents au germe et d'en sélectionner pour l'agrégation celui qui présente le coefficient de corrélation le plus proche de la valeur 1.

2.6.2 TESTS SUR LES RANGS

2.6.2.1 Test de Wilcoxon

Pour vérifier si deux échantillons indépendants proviennent de la même distribution dont les caractéristiques sont inconnues, on utilise ce test non paramétrique pour lequel on attribue à chaque observation son rang (= sa place) dans une liste ordonnée correspondant à la réunion des deux échantillons. On calcule le score W en additionnant tous les rangs du plus petit échantillon, et on le compare également aux valeurs des tables. Si min(n;m) ≥ 6, on effectue un test de Student sur les rangs, c.-à-d. un score t_0 à comparer avec la valeur prise par la distribution normale pour (n+m-2) degrés de liberté, au seuil de confiance (1-α). On rejette l'hypothèse nulle H_0 : les distributions sont identiques si $|t_0| \geq t_{n+m-2}(1-\alpha)$.
Dans ce cas également, le test n'est pas très utile pour notre problématique puisqu'il tend à éliminer les voisins les moins semblables au lieu de garder le plus ressemblant.

2.6.2.2 Corrélation sur les rangs

Ce test mesure la corrélation sur les rangs et s'exprime par le coefficient de Spearman r_s à comparer avec les tables correspondantes. Les observations de chaque échantillon sont remplacées par leur rang dans leur propre liste. Par exemple, les listes {1,0,5,8} et {6,3,2,9} deviennent {2,1,3,4} et {3,2,1,4} respectivement. Il est nécessaire de disposer de deux échantillons de même taille, car le calcul du score compare les valeurs par paire. L'hypothèse nulle de non corrélation linéaire des deux échantillons est rejetée si $|r_s| \geq r_n(1-\alpha)$.

$$r_s = 1 - \frac{6 \cdot \sum (G_i - V_i)^2}{n \cdot (n^2 - 1)} \quad \begin{array}{l} \textit{avec } G_i \textit{ et } V_i : \textit{les rangs du germe et du voisin} \\ n : \textit{taille de chaque échantillon} \end{array}$$

Équation 3 : Calcul du coefficient de Spearman : corrélation sur les rangs

Cette méthode peut se révéler utile, puisqu'elle quantifie le degré de similitude entre les échantillons, ce qui permet de hiérarchiser les voisins adjacents au germe et d'en sélectionner pour l'agrégation celui qui présente le coefficient de corrélation le plus élevé.

2.6.3 ANALYSE EN COMPOSANTES PRINCIPALES ACP

Cette méthode définit un autre système de coordonnées pour caractériser les communes concernées. Au lieu de décrire la commune par les 8 classes de production dans notre situation, l'ACP définit autant de nouvelles variables indépendantes pour l'expression des particularités de la commune. Parmi ce nouveau jeu de paramètres, on n'en conserve qu'un, deux ou trois de manière à simplifier le problème et à pouvoir représenter graphiquement les caractéristiques des communes concernées, généralement pour déduire quelles combinaisons de variables parmi les 8 classes originales comportent le maximum d'informations.

L'analyse en composante principale nécessite l'approximation de la variance s_k^2 et la covariance s_{kj} entre les variables, pour construire Σ, dite matrice de variance-covariance. Ensuite, il faut déterminer le vecteur propre ν de Σ ayant la plus grande valeur propre λ. La combinaison linéaire des variables centrées (x_i – moy(x_i)) déterminée par ce vecteur forme alors la première composante principale c_1. Pourtant, il faut d'abord effectuer un premier test en calculant les valeurs propres de Σ de manière à déterminer s'il est utile d'appliquer une telle méthode dans notre situation. Si la première valeur propre λ_1 représente plus de 60% de

la somme des λ_i (ou les trois premières valeurs propres, plus de 85%, il est envisageable de poursuivre le développement pour calculer les composantes principales de chaque commune. Malheureusement, les fonctions offertes par le langage Avenue ne permettent pas de résoudre une équation du 8e degré (8 classes de production) servant au calcul des valeurs propres, et sa programmation dépasse allègrement nos capacités ainsi que le cadre de ce travail de diplôme.

Pourtant, nous pensons utiliser différemment les résultats de cette méthode. Celle-ci ne nous offre pas réellement une alternative pour la procédure d'agrégation. Elle propose un changement de référentiel mettant peut-être à profit l'une ou l'autre méthode évincée, ne donnant pas les résultats espérés à partir des informations initiales. Nous réalisons immédiatement cette transformation sur les variables caractérisant les communes, à savoir le nombre d'entreprises de chaque spécialisation. De même, nous appliquons ce procédé sur les données 2000 de MBS primaires, conversion de variables que nous utilisons dans le chapitre 9 si elle se révèle efficace.

3 INFORMATIONS PRÉLIMINAIRES

3.1 OFFICE FÉDÉRAL DE LA STATISTIQUE

3.1.1 FONCTIONNEMENT GENERAL

La statistique agricole se base sur plusieurs textes législatifs :
- Constitution fédérale (Cst), avril 1999, article 65
- Loi fédérale de l'Agriculture (LAgr), avril 1998, article 185
- Loi sur la Statistique Fédérale (LSF), octobre 1992 et ses 4 ordonnances d'application de juin 1993
- Ordonnance sur le relevé et le traitement de données agricoles ou ordonnance sur les données agricoles (ODA), décembre 1998
- Loi fédérale sur la Protection des Données (LPD), juin 1992
- Ordonnance relative à la loi fédérale sur la protection des données (OLPD), juin 1993

Dans le cadre de ce travail, nous nous référons en particulier à la loi sur la protection des données ainsi qu'aux articles sur les restrictions de diffusion de données statistiques dans la loi sur la statistique fédérale (art. 18) et dans l'ordonnance sur les données agricoles (art. 16).

L'essentiel des nombreuses statistiques sur les exploitations, qui sont recueillies par l'OFS de manière directe ou indirecte, fait référence aux annexes de l'ordonnance sur les relevés statistiques. Toutes ces données font l'objet de traitements divers tant au niveau cantonal que fédéral. Outre ces informations sur la conduite de l'entreprise et sur ses installations, chaque centre d'exploitation fait l'objet d'un géocodage des bâtiments (= coordonnées des constructions dans le système national suisse MN03) dès février 2002.

3.1.2 DANS LE CADRE DE CE PROJET

L'OFS produit différentes cartes sur la Suisse ou sur un canton donné, représentant une ou plusieurs variables de la statistique agricole. L'agrégation manuelle des communes pour la représentation cartographique constitue un travail coûteux en temps et en personnel. Une première méthodologie est établie par Romain Tornay pour l'agrégation semi automatisée de communes dans le but de diffuser des données agricoles. Les différents critères de comparaison de commune ont été déterminés d'un commun accord entre les représentants de l'OFS et de l'EPFL. De plus, un prototype d'agrégation de communes est réalisé au sein de la chaire de SIRS, avec des hypothèses simplificatrices et la prise en compte d'un seul critère, la superficie des polygones adjacents.

3.2 DONNÉES À DISPOSITION

L'EPFL, en particulier la chaire de SIRS de la section SIE, met à disposition :
- le modèle numérique d'altitude MNA 100 sur toute la Suisse, avec une résolution de 100m qui a servi à établir un masque d'altitude
- le contour des communes et les attributs essentiels de chaque entité spatiale pour l'état 1996 (n° et nom de la commune, n° du district, n° du canton) : l'annexe IV.1 contient le détail de la correspondance entre le numéro et le nom du canton

- la version 3.2 du logiciel de SIG ArcView de la firme ESRI dont le langage de programmation choisi est Avenue
- les scripts du prototype d'agrégation

Concernant la programmation, il est également possible de télécharger et modifier des scripts mis à disposition sur le site Internet de ESRI.

L'OFS nous fournit :

- les données du recensement 1996 des entreprises du secteur primaire, en particulier celles ayant trait à l'orientation technico-économique des exploitations, à savoir :
    - Identification des communes
    - Nombre d'exploitations par commune
    - Typologie de l'orientation de production de la majorité des exploitations
    - Nombre d'entreprises spécialisées dans chaque sous-type de production
- les données du recensement 2000 sur toutes les exploitations :
    - Identification des exploitations
    - Marges brutes standard (MBS), cultures et surfaces, animaux
    - Spécialisation (classe de production)
- le contour des communes politiques (sous forme de polygones simplifiés utilisés par l'OFS pour la représentation de cartes thématiques standard)
- les attributs essentiels de chaque entité spatiale pour l'état 2000 (n° et nom de la commune, n° du district, n° du canton)

3.3 FORMAT DES DONNÉES

Description des données	**Format**
Modèle numérique d'altitude	DXF
Représentation des communes 1996 Représentation des communes 2000	Shape .shp
Nom, identifiant, n° du canton et n° du district d'origine des communes (96 et 00)	Excel .xls
Projet du prototype et scripts associés	Projet .apr scripts .ave
Nombre d'exploitation par commune 96 Typologie de l'orientation de production globale de la commune 96 Nombre d'exploitations par classe de production dans la commune 96	Excel .xls
Marges brutes standard par exploitation 2000 Surfaces (SAU, terres ouvertes, surfaces herbagères) 2000 Animaux (# têtes) 2000	Texte structuré .txt

Tableau 3 : Récapitulatif des données à disposition (description et format)

3.4 BESOINS

L'OFS Agr classe ses clients en 3 catégories :

1. le grand public, la presse et les médias "généraux"
2. les spécialistes des analyses et évaluations statistiques, i.e. les universités, les bureaux spécialisés, les organismes concernés au niveau de la recherche (Institut d'économie rurale…)
3. les chambres fédérales (commissions…) et les personnes ou organismes concernés au niveau politique (cantons, presse agricole spécialisée, étudiants…)

Les premiers clients recherchent tout d'abord une information, n'ayant pas pour but d'analyser les données ou de les utiliser pour appuyer une décision. Ils se contentent le plus souvent de graphiques et tableaux, ne demandant que peu souvent des données sous la forme cartographique. Par contre, la troisième catégorie de "consommateurs" de statistiques utilise essentiellement des cartes, pour illustrer et renforcer un argument ou comme document servant de base à une réflexion ou une conclusion. Les spécialistes commandent les deux types de données pour mieux étudier et comprendre certains phénomènes dont ils s'occupent.

Dans le domaine de l'agriculture, la statistique fédérale revêt le caractère d'une statistique transversale qui touche et intègre de nombreuses sources d'informations (statistiques de la superficie, comptes économiques, statistiques des emplois, des prix…). En ce sens, la publication "Reflets de l'Agriculture Suisse" tente de donner chaque année une vue synthétique de quelques aspects de l'agriculture.

Sur le plan international, l'OFS Agr participe aux travaux de différentes organisations :

- Eurostat, dans plusieurs groupes de travail traitant des concepts et des programmes de production statistique, ainsi que de l'analyse des résultats
- OCDE, FAO, CEE
- Office nationaux de statistique, avec lesquels une étroite collaboration s'est installée dans les domaines du traitement des données des DataWarehouses.

Dans chacune de ces relations, l'OFS Agr est amené à présenter des informations sous forme cartographique, et ceci parfois dans des délais relativement courts. Pour un meilleur traitement des commandes et pour une efficacité accrue au sein de la section, l'intérêt se porte sur des procédures d'automatisation dans l'élaboration des produits cartographiques. C'est dans ce but ultime que nous développons un prototype pour l'agrégation d'entités territoriales pour permettre à l'OFS de diffuser plus facilement ses informations. En effet, l'élaboration d'un logiciel complet dépasse le cadre réservé à un travail de diplôme de durée limitée. Par contre, il nous semble important d'évoquer les apports et les limites de la technologie actuelle offerte par les SIRS, en rapport avec la problématique de ce travail. Nous étendrons nos réflexions à travers des perspectives de production et d'utilisation de telles techniques au sein de l'OFS, sans donner de solutions définitives, mais plutôt quelques indices pour résoudre de nouvelles situations.

4 CRITÈRES D'AGRÉGATION

Comme mentionnés auparavant, la décision et le choix des agrégats s'effectuent encore par voie humaine, sur la base de plusieurs facteurs, de manière à assurer la cohérence et l'esthétisme du résultat. Dans le cadre de ce travail de diplôme, le choix des critères d'agrégation s'est réduit à reprendre les critères établis tout d'abord par l'OFS et rediscutés avec R. Tornay lors de son travail de fin de cycle postgrade à l'EPFL.
Nous avons obtenu l'explication et la justification de chaque critère de manière à accorder nos points de vue sur l'orientation à donner à ce projet. Nous avons donc choisi de suivre les mêmes principes directeurs en gardant le même jeu de critères, relativement restreint puisqu'il ne comporte que 4 éléments, mais suffisamment étoffé pour représenter tous les intérêts entrant en ligne de compte. Il s'agit alors des catégories suivantes :

1. Confidentialité : soumis à la loi sur la protection des données, l'OFS ne peut diffuser ses données que sous une forme anonyme
2. Ressemblance : pour que la carte représente le plus fidèlement possible les données récoltées, il est préférable de regrouper des communes ayant des caractéristiques semblables -si ce n'est identiques- concernant le thème considéré.
3. Appartenance au canton : comme un grand nombre de statistiques sont également cartographiées à l'échelle cantonale, nous souhaitons favoriser l'agrégation entre communes d'un même canton.
4. Esthétisme : par soucis de lisibilité de la carte, nous voulons obtenir des polygones de forme régulière.

Un premier développement sur les données 96 aboutit à l'élaboration d'un prototype reflétant les orientations initiales définies lors du travail de R. Tornay. Ce programme se base sur les quatre critères et sept indicateurs suivants :

Critères	Indicateurs
Confidentialité	Nombre d'exploitations agricoles
Ressemblance	Nombre d'exploitations agricoles Typologie de l'orientation de production
Appartenance	Canton d'origine
Esthétisme	Indice de compacité (Gravélius) Proportion de la frontière en commun Taille de l'agrégat (# d'entreprises)

Tableau 4 : Présentation des critères et indicateurs de la méthode multicritère appliquée aux données de 1996

A noter que l'indicateur de confidentialité n'intervient pas directement dans le processus d'agrégation puisqu'il sert à sélectionner les germes, ainsi qu'à assurer la pertinence des agrégats.

Le choix des indicateurs et leur justification sont présentés ci-après.

4.1 CRITÈRE DE CONFIDENTIALITÉ

Il s'agit d'une des raisons principales ayant amorcé le lancement de ce projet. Rappelons en effet que, pour diffuser des résultats, l'OFS Agr se voit contraint de rassembler les données de plusieurs communes pour garantir l'anonymat des exploitations concernées. En terme de représentation cartographique, ce regroupement correspond à l'agrégation des polygones symbolisant les communes considérées.
Nous choisissons le nombre d'exploitations présentes sur le territoire communal comme unique indicateur pour ce critère. De plus, pour respecter les contraintes liées à la protection des données, nous utilisons le seuil de confidentialité établi par l'OFS pour les informations sur l'orientation de production et fixé à 12 exploitations au minimum par agrégat. Cette valeur est plus ou moins arbitraire et doit pouvoir être modifiée selon les circonstances (objectifs de l'agrégation, nature de la variable à cartographier…). Elle est issue de l'expérience acquise à l'OFS pour la cartographie et réduit à environ 2000 le nombre de communes suisses, ce qui est raisonnable pour les manipulations et pour une vision globale et claire de la Suisse. En effet, selon une convention respectée par l'OFS ainsi qu'au niveau international, la confidentialité est respectée lorsque 4 exploitations au moins par communes remplissent le critère sur lequel est jugée la ressemblance. Pour garantir en toutes situations l'anonymat des données d'orientation de production cartographiées sous la forme d'une typologie, la valeur de 12 exploitations constitue un fort gage de sécurité.

L'indicateur de confidentialité est directement caractérisé par le nombre total d'exploitations de chaque commune répertorié comme attribut à part entière dans la base de données sur les communes (visible sous forme de tableau dans ArcView, par après "table des communes").

4.2 CRITÈRE DE RESSEMBLANCE

Le critère de ressemblance se définit par deux indicateurs : le nombre d'exploitations de chaque entité et l'objet de la cartographie, en l'occurrence l'orientation de production.

Mentionnons que le second indicateur est le seul véritable paramètre qui est amené à varier lors de la phase opérationnelle du projet. Le prototype est développé à partir des données concernant l'orientation de production, mais son application devrait être possible pour la représentation cartographique d'autres thèmes privilégiés par la section d'agriculture, voire par l'OFS en totalité.
C'est pourquoi il est préférable que l'information concernant cet indicateur soit directement disponible dans la table des communes.

Pour le cas de l'orientation de production, neuf champs en donnent les caractéristiques principales : le nombre d'exploitations spécialisées dans chacune des huit classes de production, et le résumé sous la forme d'un code, la typologie. (cf. annexe II.1)

4.3 CRITÈRE D'APPARTENANCE AU CANTON

Lors de la première phase de travail sur ce sujet, les parties s'accordent pour établir un critère d'appartenance, avec le canton comme indicateur. Il s'agit de limiter l'agrégation d'une commune à une autre du même canton. En effet, de nombreuses statistiques sont rassemblées et publiées à l'échelle cantonale. Les offices cantonaux de statistiques récoltent bon nombre de données dans tous les domaines de la statistique, et en particulier sur la démographie, l'économie et l'agriculture. Ils fournissent leurs relevés à l'OFS qui leur en procure d'autres ou leur restitue leurs informations sous d'autres formes.
Lors de cette étape, un petit doute s'installe déjà sur la signification et l'importance de ce paramètre. Malgré tout, une ou deux alternatives sont d'ores et déjà proposées, à savoir la possibilité d'effectuer l'agrégation un canton après l'autre. D'autre part, pour une procédure d'agrégation globale sur tout le territoire suisse, une solution consiste à attribuer un poids moyen ou faible au critère d'appartenance lors de la résolution multicritère. Ainsi, les délimitations cantonales sont respectées dans la plupart des cas.

Quoi qu'il en soit, l'information concernant le canton d'origine est rapidement accessible dans la table des communes par le numéro du canton. (cf. annexe IV.1)

4.4 CRITÈRE D'ESTHÉTISME

Trois indicateurs ont été retenus pour ce critère :
1. Un indice de compacité
2. La frontière commune
3. La taille de l'agrégat

4.4.1 L'INDICE DE COMPACITE

Il existe de nombreuses manières d'apprécier la forme d'une entité surfacique, chaque domaine d'activité utilisant les indices les plus appropriés (indices de compacité, de circularité…). Nous empruntons un de ces indicateurs à un domaine s'appuyant régulièrement sur les technologies des SIRS, comme l'hydrologie par exemple, de manière à vérifier si l'agrégat présente une forme suffisamment régulière pour garantir une vision claire de la carte.

Parmi les indices à disposition, nous avons sélectionné trois indices de compacité (P. Haggett, 1977 et A. Musy, 2001):

- L'indice de Gibbs (1961): $K_G = \frac{1.2373 \cdot S}{L^2}$ **Équation 4 : indice de compacité de Gibbs**
- L'indice de Cole (1964): $K_C = \frac{4 \cdot S}{pL^2}$ **Équation 5 : indice de compacité de Cole**
- L'indice de Gravélius (1914) : $G = \frac{P}{2\sqrt{pS}}$ **Équation 6 : indice de compacité de Gravélius**

Avec S représentant la surface du polygone, P son périmètre et L son grand axe (= distance entre les deux points les plus éloignés de l'unité, donnant le diamètre du cercle circonscrit.).

Les deux premiers sont utilisés principalement par des géographes, alors que le dernier convient bien à l'étude de bassins versants en hydrologie. On remarque tout de suite qu'ils se ressemblent tous puisqu'ils font intervenir plus ou moins les mêmes paramètres. Nous pouvons légitimement supposer que, dans le cas présent et pour le but que nous poursuivons, tous ces indices pourraient être utilisés et produire des résultats semblables.

Par sa formulation, il est aisé de se représenter géométriquement la signification de l'indice de Gravélius puisqu'il compare la forme considérée avec un cercle. En effet, il est défini comme le rapport du périmètre du polygone au périmètre du cercle ayant la même surface. De plus, les composantes de l'équation peuvent facilement être déterminés à partir de la géométrie définissant chaque polygone dans ArcView. Nous décidons alors d'adopter cet indice pour caractériser la forme d'une commune ou d'un possible agrégat entre deux communes, d'autant plus qu'il est connu autant à l'EPFL qu'à l'OFS.

Tant le périmètre (P) que la surface (S) peuvent être calculés à partir de la géométrie des polygones stockée dans la table des communes.

Le schéma ci-dessous met en évidence quelques valeurs prises par l'indice, exemple appliqué au domaine de l'hydrologie :

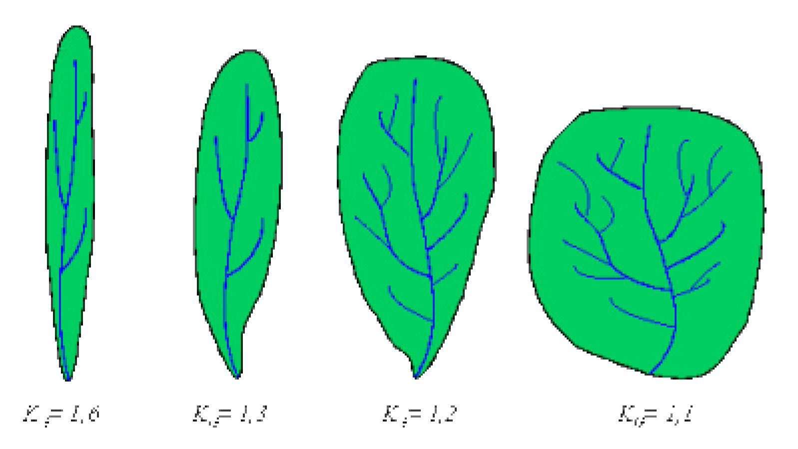


Figure 2 : Application de l'indice de Gravélius à l'étude de bassins versants en hydrologie
Source : SSIE, Institut d'Aménagement des Terres et des Eaux (IATE), Laboratoire d'Hydrologie et Aménagement (HYDRAM)

4.4.2 LA FRONTIERE COMMUNE

La frontière commune est un facteur simple dans sa conception et relativement efficace bien qu'intuitif. En effet, il paraît logique que l'agrégation d'un germe avec le voisin qui présente la plus grande frontière commune aboutisse à un polygone de forme plus régulière que si les communes de base ne partagent qu'une faible part de leurs limites. Cet indicateur est utilisé pour renforcer l'action de l'indice de Gravélius en minimisant leur impact respectif sur le critère de ressemblance.
Si nous désirons appliquer une résolution multicritère, nous pensons qu'il est utile de normaliser cet indicateur, de manière à faciliter l'attribution d'un score. Nous proposons de rapporter la longueur de la frontière commune au périmètre du germe, pour ne manipuler que des valeurs relatives, en [%].

Pour le calcul de la frontière commune, nous utilisons deux fonctions spécifiques de ArcView nous permettant d'extraire la géométrie de ce fragment, par intersection des deux périmètres des polygones concernés, et d'en calculer la longueur par la suite.

4.4.3 LA TAILLE DE L'AGREGAT

L'importance de la taille de l'agrégat se concrétise principalement lors de l'agrégation des communes tessinoises. Comme la majorité de ces entités sont des germes, il faut éviter que le regroupement ressemble à un effet boule de neige autour des quelques communes importantes. Dans un tel cas de figure, on se retrouverait avec une situation inacceptable où seulement une poignée d'agrégats couvriraient le Tessin. La nécessité devient alors évidente d'établir une limite quant à la taille des agrégats.

Deux paramètres possibles viennent immédiatement à l'esprit : la surface ou le nombre d'exploitations de l'agrégat. On peut discuter de l'efficacité de l'un ou et l'autre, mais il semble qu'aucun des deux ne s'impose particulièrement. Le nombre d'exploitations est choisi comme complément au critère de confidentialité, c'est-à-dire qu'il fixe une limite supérieure à l'agrégat au lieu de requérir un minimum.

Dans ce cas également, l'information est directement accessible dans la base de données relative aux communes.

5 PRINCIPALES ÉTAPES D'AGRÉGATION

Pour cette première élaboration d'un programme d'agrégation, nous nous sommes basés sur les données 96, soit un contour des communes fourni par l'EPFL et les données (typologie et spécialisation des communes) du recensement 96 de l'OFS Agr.

5.1 DESCRIPTION THÉORIQUE

Tout d'abord, un masque d'altitude a été créé d'après le modèle numérique de terrain MNA 100 d'une résolution de 100m, en sélectionnant tous les points supérieurs à 1600m d'une part et à 2000m dans une seconde application. Toute la phase de transformation d'une grille (sélection de pixels du MNA) vers un polygone (contour vectoriel en format Shape .shp) a été réalisée par un collaborateur de la chaire de SIRS de l'EPFL.

Ensuite, le découpage des communes suivant ce masque a fragmenté certaines communes. Cette complication s'est greffée au problème des communes représentées par un polygone complexe, entité composée de plusieurs formes géométriques non adjacentes. La dissociation de ces polygones, l'allocation des attributs au plus grand fragment et l'agrégation des petites sections à leur plus grand voisin en surface ont toutes été programmées par ce même collaborateur et testées avec le masque posé à 1600m d'altitude.

L'étude de ces scripts nous a permis de nous familiariser avec le langage Avenue et de comprendre les différents obstacles liés à la programmation et à la configuration géographique des polygones. Nous avons utilisé le masque à 2000m (format vectoriel) pour effectuer nos propres tests à l'aide de ces scripts, afin d'effectuer le parallèle entre les fonctions programmées et les opérations réalisées sur les polygones et sur les données tabulaires.

Enfin, nous avons pu entamer la phase d'agrégation proprement dite, en reprenant les étapes du prototype à disposition concernant la réunion physique et tabulaire des communes. Le code définissant la sélection des germes et la désignation du voisin le plus ressemblant a été remodelé pour construire une procédure d'agrégation selon un modèle multicritère.

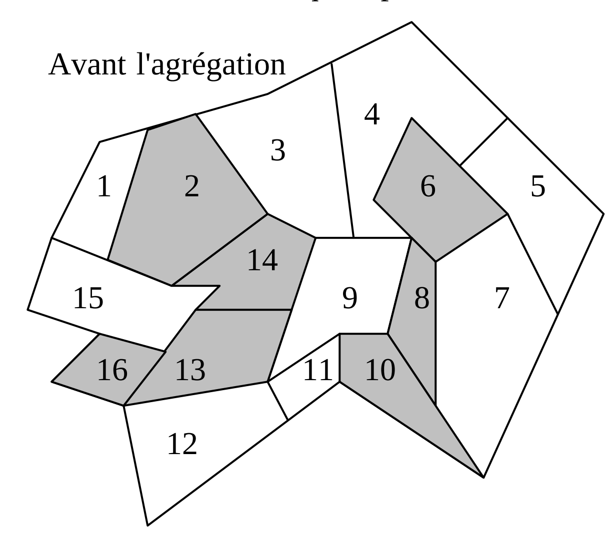
Ainsi, le processus d'agrégation suit les principales étapes suivantes :

1. Sélectionner les germes d'après le nombre d'entreprises agricoles des communes
2. Pour chaque germe, sélectionner les voisins adjacents et trouver le meilleur (pour la méthode multicritère, se référer au paragraphe 2.5)
   a. pour chaque voisin, attribuer un score ou un indice
   b. déterminer le voisin le plus ressemblant (score maximal), éventuellement départager les candidats ex aequo sur la base d'un champ discriminant (choisi par l'opérateur)
   c. insérer les n° identifiants du germe et du voisin dans deux listes distinctes
3. Epurer les listes (cf. schéma explicatif ci-dessous : figure 3)
   a. ôter tous les germes (sauf 1) voulant s'agréger à la même commune
   b.1 éliminer une des deux relations où une commune se retrouve en même temps agrégeante et agrégée
   b.2 sauver de cette sélection les communes voulant s'agréger réciproquement

4. Effectuer l'agrégation proprement dite (cf. figure 3)
    a. pour chaque commune, enregistrer les attributs
    b. agréger physiquement : fusion de polygones
    c. définir les attributs de la nouvelle commune
        i. sommer des attributs des deux communes initiales (aire, nombre d'exploitations par commune et par spécification)
        ii. reprendre des attributs de la commune agrégeante (nom, canton…)
        iii. recalculer des attributs (typologie, état : germe ou entité…)
5. Calculer le nombre de germes non encore agrégés et effectuer les itérations nécessaires

5.2 SIMULATION D'AGRÉGATION

Dans l'exemple ci-contre, les germes sont les communes n° 2-6-8-10-13-14-16, et toutes les différentes situations qu'on peut rencontrer sont représentées



Etape 1 : La commune n° 16 est écartée directement car elle ne contient aucune exploitation. (*)

Etape 2 :Au terme de la sélection des meilleurs voisins, les couples suivants se sont formés :
(2→1) (6→8) (8→7)
(10→8) (13→14) (14→13)

Etape 3 : Ensuite, on élimine toutes (sauf 1) relations où plusieurs communes veulent s'agréger à la même commune. On obtient
(2→1) (6→8) (8→7)
(13→14) (14→13)

Etape 4 : On construit une liste des relations où la commune agrégeante est également à agréger dans une autre relation, soit dans cet exemple, les couples
(6→8) (13→14) (14→13)

Etape 5 : De cette sélection, on gracie un couple dans le cas où des communes veulent s'agréger mutuellement, soit ici le coupe (14→13) et il nous reste, dans la liste précédente, les relations (6→8) et (13→14) à éliminer de la liste issue de l'étape 3

Etape 6 : On effectue l'agrégation des polygones et de leurs attributs. Dans le cas présent, il reste à effectuer réellement l'agrégation sur les couples
(2→1) (8→7) (14→13)

Figure 3 : Exemple théorique d'agrégation

(*) Remarque : les communes sans exploitation ne sont pas traitées dans la procédure comme des germes, mais agrégées en dernier ressort. En effet, il est impossible d'établir une ressemblance significative avec ces communes. De plus, il est préférable d'effectuer l'agrégation à la fin, le seul critère étant d'ordre esthétique.

Tous les tests réalisés avec les données 96 de l'OFS Agr se basent sur la représentation initiale des communes 96 à laquelle nous avons appliqué le masque d'altitude et la procédure d'élimination des communes complexes, à savoir la conservation du plus grand fragment d'une commune composée initialement de plusieurs polygones, et l'agrégation des autres parties à leur plus grand voisin (cf. annexes IX.1 et 2). De plus, quelques tests sur un programme multicritère sont effectués en premier lieu afin d'établir un jeu de poids permettant d'aboutir à un nombre d'agrégations assez élevé, pour un résultat satisfaisant. Nous mettons un accent particulier sur cette méthode car elle se rapproche le plus de la procédure manuelle appliquée jusqu'à présent. Elle servira de base de comparaison face à d'autres procédés envisagés, comme les techniques statistiques telles que la corrélation. Un résumé de l'échelle de scores et des poids attribués aux différents indicateurs est exposé en annexe V.1.

Le détail des essais n'est pas présenté dans ce rapport : n'ayant comme référence que le nombre d'agrégats espérés et une carte (papier) des orientations de production sur les communes originales, nous avons effectué des tests visuels en premier lieu. De plus, nous avons choisi le nombre total d'agrégats et le nombre de communes agrégées, respectivement agrégeantes, conservant leur orientation de production lors de l'agrégation pour juger de l'efficacité de la méthode. Nous n'avons eu besoin que de trois tentatives pour trouver un jeu de poids donnant des résultats satisfaisants. L'objectif étant de rapidement progresser et migrer sur les données 2000, nous n'avons conservé que les résultats de la solution retenue.

D'autre part, la progression de la valeur limite du nombre d'exploitations (VL) pour la sélection des germes est une notion apparue très tôt dans les discussions et apportant rapidement les résultats escomptés. Cette solution s'est de suite imposée comme référence, tout en gardant en réserve la possibilité d'utiliser une VL fixe sur l'ensemble des itérations nécessaires à l'élimination des germes.

Pour toutes les prochaines informations des chapitres 5 à 7 contenant d'une séquence de poids, l'ordre des valeurs correspond à la liste des indicateurs ci-dessous :

1. Appartenance au canton
2. Ressemblance : nombre d'exploitations
3. Ressemblance : orientation de production
4. Esthétisme : frontière commune
5. Esthétisme : indice de Gravélius
6. Esthétisme : taille de l'agrégat

Par exemple, le jeu de poids ayant abouti à la meilleure agrégation jusqu'à ce point se présente sous la forme (1-3-5-2-2-1), valeur de référence.

6 SITUATIONS CONTENTIEUSES

Dans ce chapitre sont présentés les résultats concrets obtenus lors de l'application du programme d'agrégation par résolution multicritère. Les données de base sont les polygones représentant les communes et les informations du recensement 96 des entreprises agricoles. Nous attribuons le jeu de poids de référence (1-3-5-2-2-1) aux indicateurs selon l'ordre détaillé au chapitre précédent. Le critère de confidentialité est assuré par l'établissement d'une valeur limite minimale (VL) du nombre d'exploitations de la commune. Nous choisissons de varier cette VL (2, 4, 8, 12 et 12) lors de 5 itérations du programme, de manière à favoriser l'agrégation des "petites" communes en premier. Ci-dessous, nous exposons en images quelques situations où le résultat de l'agrégation, telle que programmée, n'est pas toujours optimale. Toutes les illustrations sont présentées avec le Nord vers le haut de la page. Nous n'y avons pas mentionné d'échelle, car elle n'est pas constante d'une carte à l'autre et n'apporte aucune information supplémentaire pour la comparaison de la forme et de la typologie des agrégats.

6.1 MAUVAISE AGRÉGATION DU POINT DE VUE DE LA FORME DE L'AGRÉGAT

Les deux images ci-dessous présentent quelques situations où les indicateurs de forme n'ont pas été suffisamment puissants. Dans les deux cas, c'est l'indicateur du nombre d'exploitations utilisé comme indicateur de ressemblance qui a surpassé la contrainte posée par l'indice de Gravélius et la longueur de la frontière commune. On peut déjà mettre en doute la nécessité d'un tel indicateur de ressemblance.

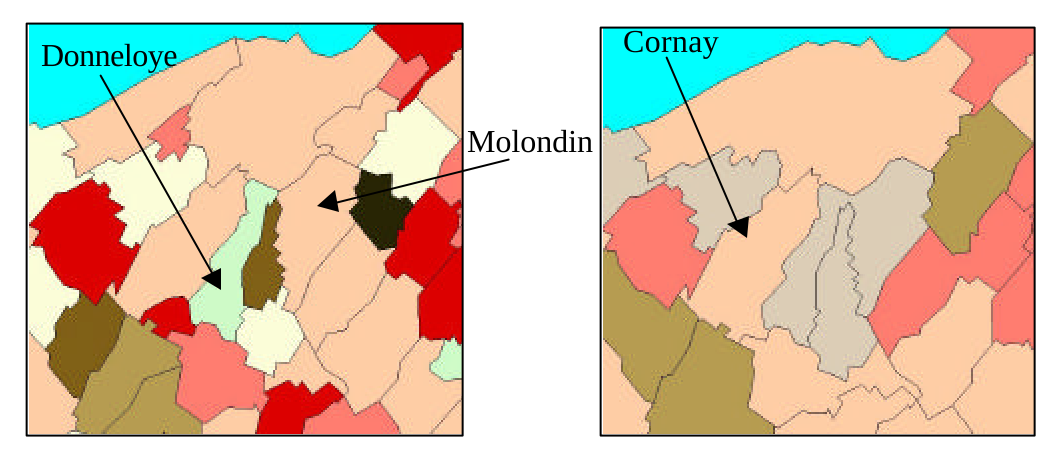
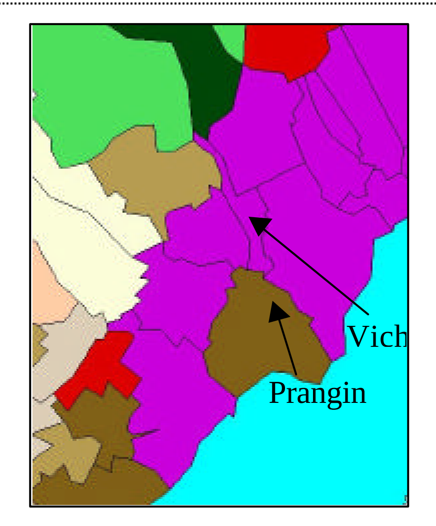


Figure 4 : Agrégat de forme irrégulière : exemple 1

Dans l'exemple ci-dessus, il aurait été plus judicieux de réunir Donneloye et Cornay au lieu de l'agrégat effectué entre Donneloye et Molondin. De même, le groupement Prangin – Vich de la figure 5 serait avantageusement remplacé par les associations Prangin – Gland et Vich – Genolier – Coinsins.



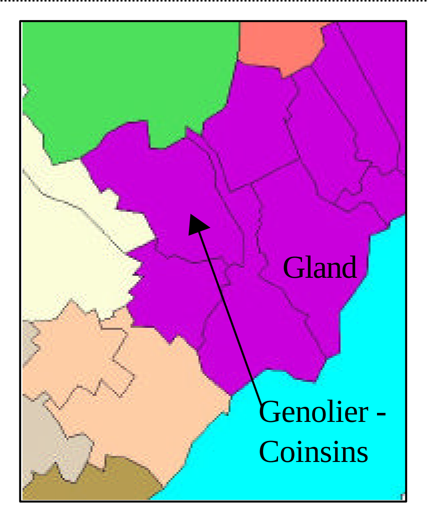
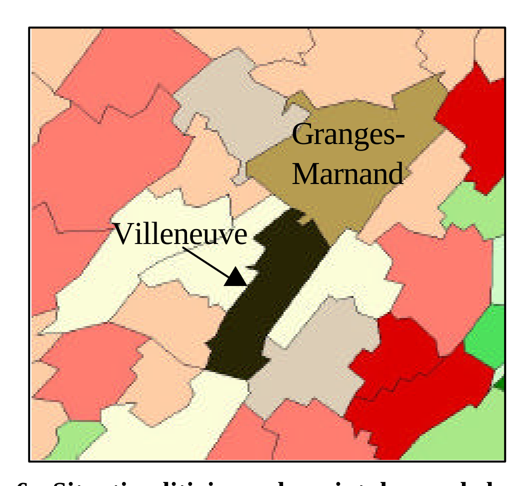


Figure 5 : Agrégat de forme irrégulière : exemple 2

6.2 MAUVAISE AGRÉGATION DU POINT DE VUE DE LA RESSEMBLANCE

Nous pouvons observer qu'il n'y a pas d'erreur flagrante lors de l'agrégation utilisant la méthode multicritère, mais juste quelques ambiguïtés. Prenons l'exemple de la commune de Villeneuve (FR) (figure 6 en brun foncé), qui, alors qu'elle est spécialisée en culture végétale à plus de 70%, elle se retrouve agrégée à des exploitations mixtes, produisant de surcroît un agrégat à dominante animale.



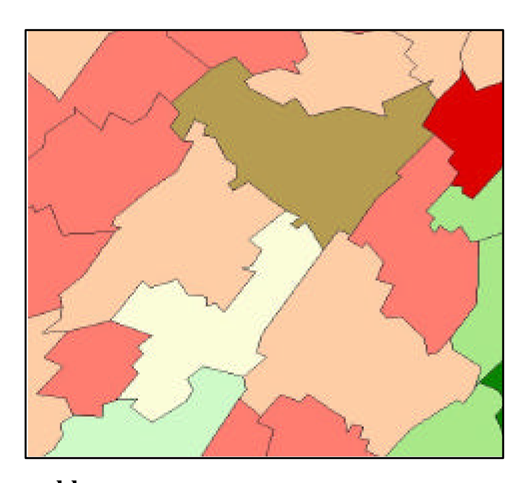


Figure 6 : Situation litigieuse du point de vue de la ressemblance

Dans cet exemple s'exprime en réalité une grande partie de la complexité de la problématique, ainsi que la difficulté de trouver des indicateurs efficaces et de leur attribuer des poids pertinents. Si l'on raisonne uniquement en terme de ressemblance, une réunion avec Granges-Marnand (VD) (en marron au Nord-NE) aurait été meilleure. En plus, l'information thématique se trouve un peu lissée, puisqu'il ne reste que des teintes pastel (spécialisation peu dominante) et une majorité d'agrégats mixtes. Une telle configuration finale relève aussi probablement de la procédure d'agrégation appliquée. En utilisant des valeurs limites successives croissantes, on force les "petites" communes à s'agréger en premier. Dans l'extrait ci-dessus, Villeneuve (10 exploitations) s'est certainement trouvée parmi les dernières communes à devoir s'agréger.

D'autre part, on peut observer que le résultat est globalement bon au niveau de la forme des agrégats. En outre, le programme suit les contraintes implicites souhaitées puisque l'agrégation entre une commune d'orientation animale et une de production végétale est évitée autant que possible. Finalement, on peut remarquer que la frontière cantonale a été respectée, diminuant encore la probabilité d'association entre Villeneuve et Granges-Marnand.

Ce dernier point permet également de justifier la réunion de Marin (NE) (en brun foncé) à Saint-Blaise (au Nord-Ouest) sur la figure 7 : c'est le critère d'appartenance au canton qui a décidé de l'agrégation finale entre ces deux solutions assez proches. En effet, la commune de Gampelen (BE) se retrouve écartée au profit de St-Blaise qui appartient au même canton que Marin, en dépit d'un léger avantage du point de vue de la ressemblance.

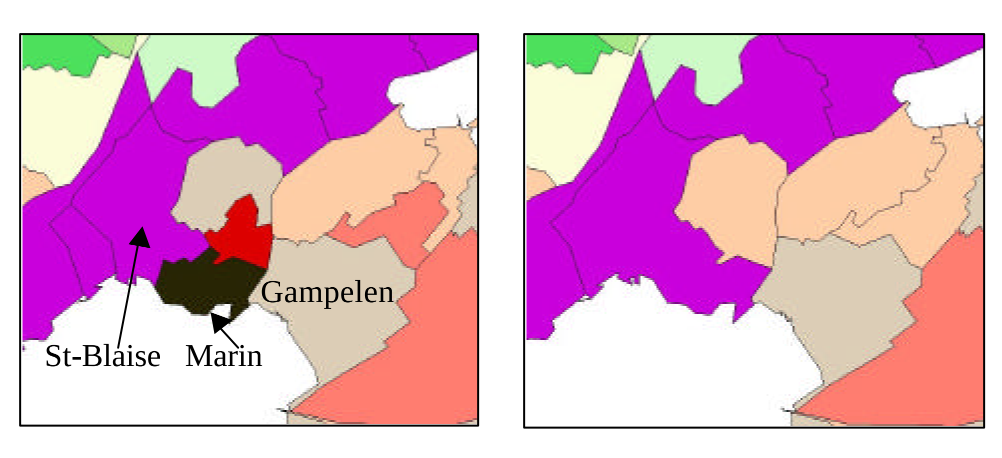


Figure 7 : Influence du critère d'appartenance au canton

Si, dans le cas de Villeneuve et de Marin, il était tout de même possible de réunir le village concerné à une commune de production semblable en modifiant le jeu de poids par exemple, ce n'est pas toujours le cas. Voici quelques exemples où même l'intervention humaine se révélerait impuissante.

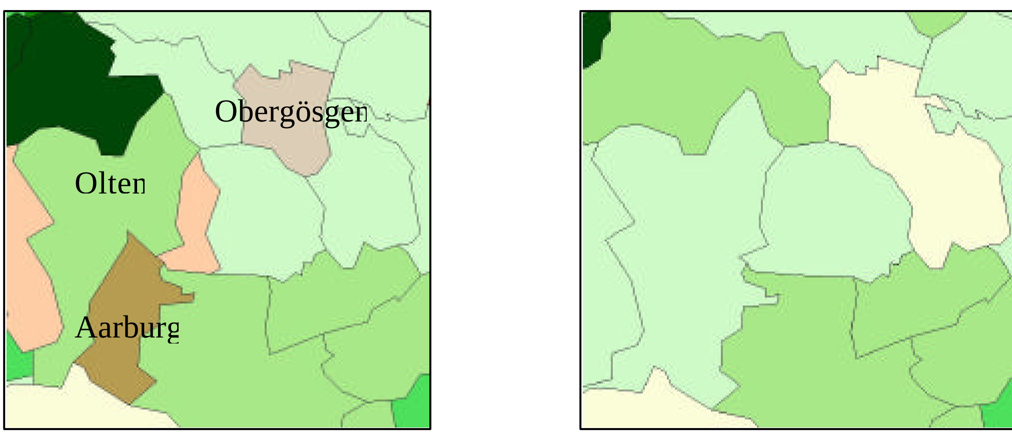


Figure 8 : Commune isolée du point de vue de la typologie : exemple 1

Dans la région d'Olten (ci-dessus) comme vers Zurich (ci-dessous), certaines communes se distinguent puisqu'elles sont entourées de communes ayant toutes une orientation de production antagoniste à celle de la commune en question. C'est le cas ici pour Aarburg, Obergösgen, Rümikon et Zollikon.

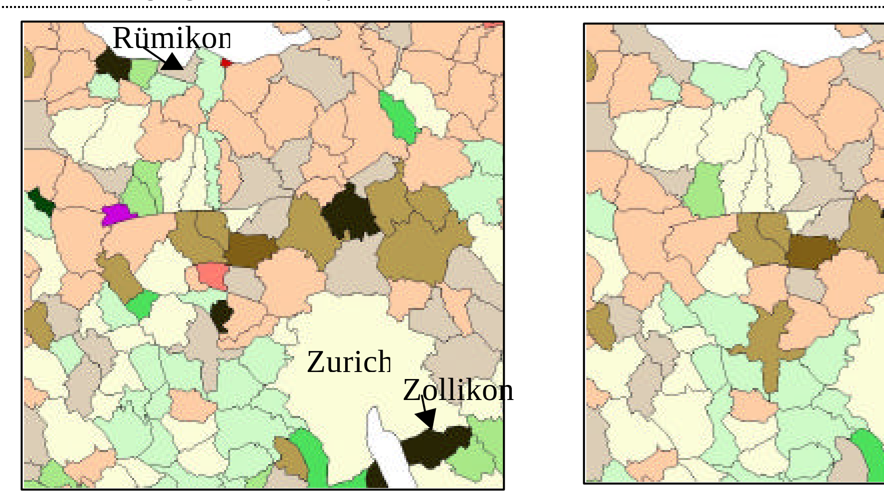


Figure 9 : Commune isolée du point de vue de la typologie : exemple 2

Même si on veut l'éviter au maximum, l'agrégation entre communes d'orientations antagonistes est rendue inéluctable dans de telles circonstances. Pour les communes mixtes, la question se pose de savoir si l'on adopte une agrégation préférentielle avec un type de production donné. En effet, il pourrait être intéressant d'orienter davantage l'agrégation, en examinant plus la structure conjointe entre les communes mixtes et leurs voisines végétales ou animales. Pourtant, au vu des biais induits par la définition d'une typologie, cette perspective ne nous semble pas judicieuse dans le cas présent. Nous préférons laisser ce choix libre pour favoriser l'esthétisme de la carte. En effet, c'est l'agrégat présentant la forme la plus régulière qui sera choisi, quelle que soit l'orientation technico-économique du voisin.

6.3 LE CAS PARTICULIER DU TESSIN

Alors que, de manière générale, les communes de Suisse alémanique engendrent peu d'agrégats litigieux, l'endroit présentant le plus de cas particuliers est sans conteste le canton tessinois, avec ces nombreuses communes ne comportant qu'une ou deux entreprises agricoles. Un exemple, tiré des bords du lac de Lugano, résume l'ensemble des configurations spéciales détaillées auparavant.

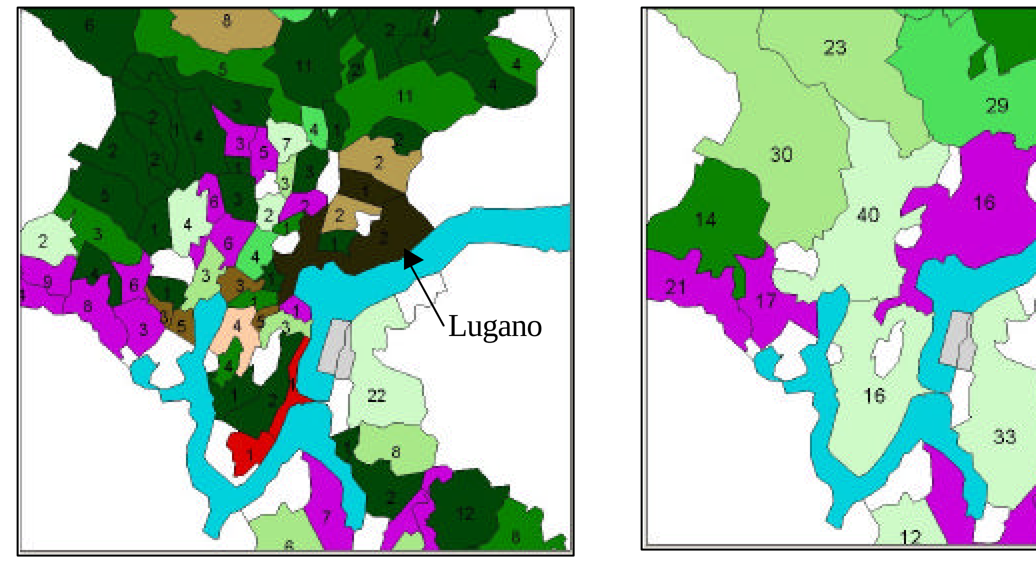
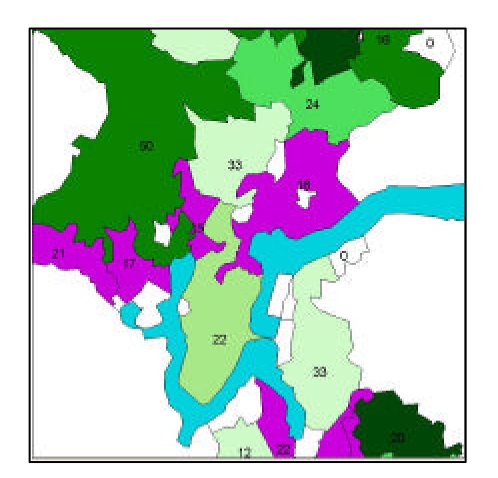


Figure 10 : Tessin, cas particulier sur les bords du lac de Lugano (situations initiale et de référence)

En surimpression se trouve le nombre d'exploitations de chaque commune : on peut remarquer qu'il est impensable de former naturellement des unités d'une seule orientation de production, d'autant plus qu'on commence par réunir les "petites" communes. Ce problème se répercute sur la forme de l'agrégat et sur sa typologie finale, quel que soit le mode d'agrégation choisi. Dans ce cas de figure, il semble pourtant envisageable de trouver une meilleure solution qu'illustrée par la figure 10, en essayant de limiter davantage la taille finale des agrégats par exemple (cf. figures 13 et 14).

Nous avons effectué quelques essais, en exécutant le nombre d'itérations nécessaire à l'élimination de tous les germes.
Nous appliquons d'abord le jeu de poids de référence (1-3-5-2-2-1) avec la valeur limite du nombre d'exploitations fixe (VL = 12), puis en échelonnant davantage les valeurs attribuées au seuil VL (VL = 2, 4, 6, 8, 10, 12)



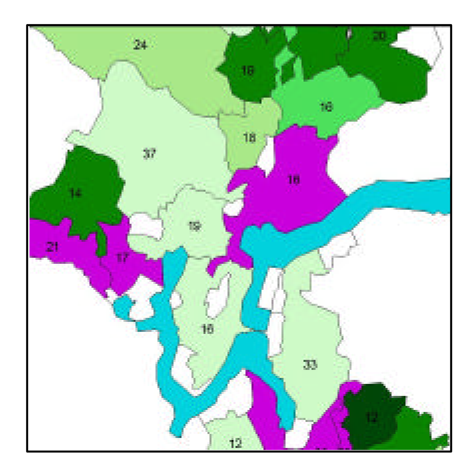


Figure 11 : Tessin : variations de la VL sur le nombre d'exploitations, avec les poids de référence

Dans les deux cas, la différence avec la solution initiale correspondant à l'application de référence n'est pas marquante. L'alternative consistant à garder un seuil fixe pour la contrainte de confidentialité peut dès lors être définitivement écartée : les agglomérats sont trop "volumineux" et peu esthétiques. Dans le second exemple, le lissage induit par la multiplication des itérations nous pousse également à ne pas retenir cette option par la suite.

Ensuite, nous testons une alternative caractérisée par une pondération uniforme (1-1-1-1-1-1), une première application avec la valeur limite du nombre d'exploitation variable VL = 2, 4, 8, et 12 successivement (image de gauche) et une seconde manipulation en gardant VL = 12 constante (image de droite).

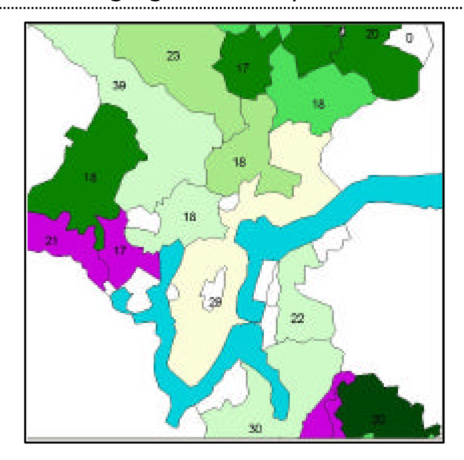
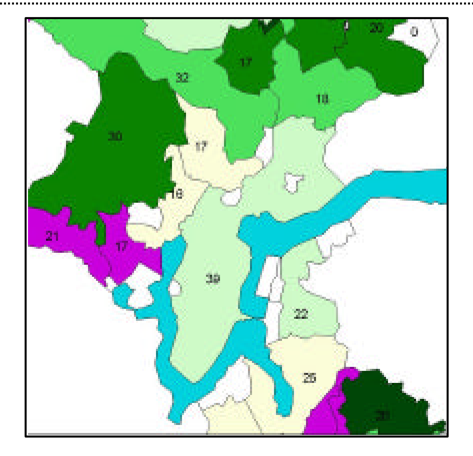



Figure 12 : Tessin : alternatives avec pondération uniforme

Les résultats représentés par ces deux situations ne sont guère favorables : les agrégats sont trop grands en surface et en nombre d'exploitations. En outre, l'information originale se retrouve complètement lissée sur les bords du lac, puisque des classes de production distinctes sont englobées dans un seul agrégat, sans orientation véritablement marquée.

D'autre part, nous avons effectué deux expériences en attribuant un poids plus fort à l'indicateur de taille (poids 1-2-4-1-1-5), selon les mêmes séquences de valeurs limites VL que ci-dessus.

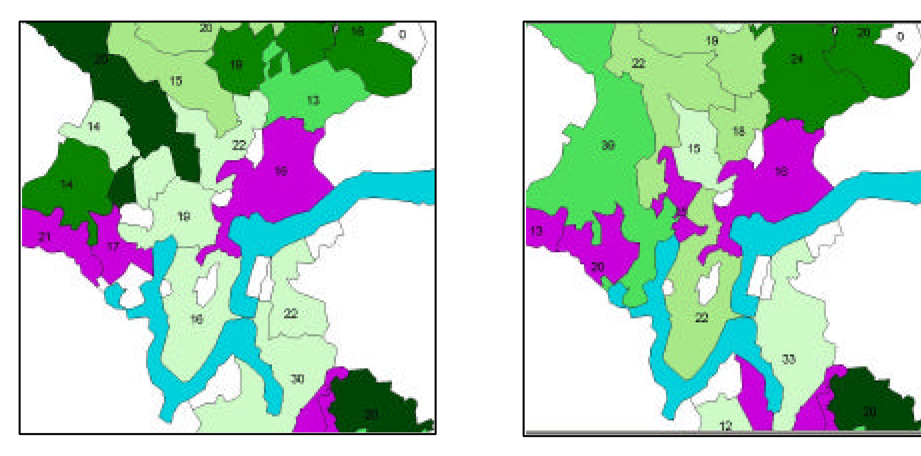


Figure 13 : Tessin : alternatives avec un poids fort sur l'indicateur de taille

L'expérience n'est pas concluante, puisque les formes des agrégats sont plus irrégulières et que leur typologie finale ne correspond que très partiellement à la répartition originale.

Finalement, nous tentons encore deux essais renforçant l'action de l'indicateur de taille. Intuitivement, il nous paraît important d'accentuer le poids du facteur "taille" de manière à conserver des agrégats proches du seuil de confidentialité garantissant une meilleure représentation des données de base. Nous appliquons les poids 1-0-1-3-3-5 et 1-0-1-1-0-5 avec une VL variable (VL = 2, 4, 8, 12) pour obtenir les résultats suivants.

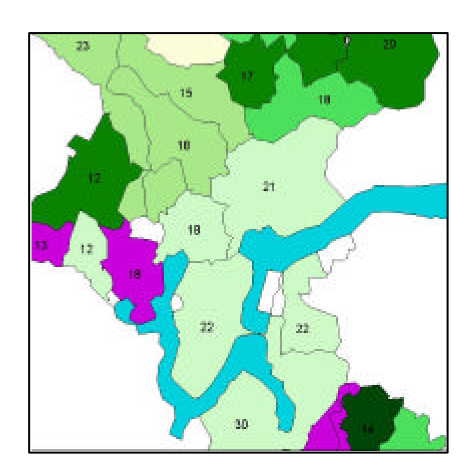
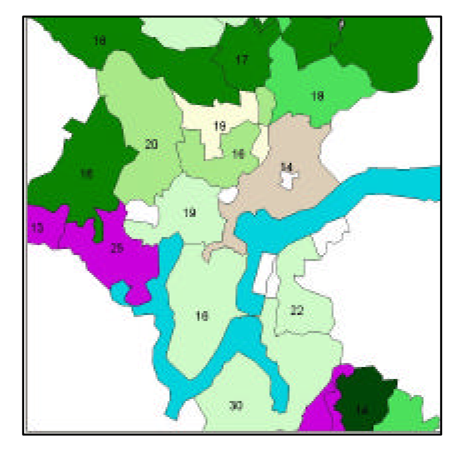



Figure 14 : Tessin : variantes avec indicateurs d'esthétisme prédominants

L'essentiel des remarques précédentes peut également s'appliquer à la première image, à l'exception de la taille et de la forme des agrégats qui sont ici plutôt uniformes et régulières.
Malgré la présence de formes trop irrégulières, le second résultat apporte des changements probants par rapport à la solution de référence. En effet, la taille des agrégats est raisonnable tant au niveau spatial que sur le nombre d'entreprises agricoles. De plus, à l'exemple des cultures végétales également représentées sous une autre forme que les cultures permanentes, la majorité des classes de production retrouve une affectation relativement fidèle à la distribution initiale.

Si l'on compare les résultats sur tout le Tessin (cf. annexes VI.1 à 3), on remarque que les observations effectuées sur cet extrait de carte ne peuvent pas être simplement généralisées à l'ensemble du territoire cantonal. En effet, chacune des solutions présente ses points forts et ses inconvénients : la résolution d'un problème sur les bords du lac de Lugano est couplée à l'apparition de formes irrégulières plus au Nord, et à l'émergence d'agrégats moins représentatifs dans d'autres parties du canton. Si l'on juge par les quelques chiffres suivants, il n'y a pas de solution idéale, mais plusieurs variantes représentant chacune un compromis favorable.

		Communes agrégées inchangées		Communes agrégeantes inchangées	
Poids	**Nombre d'agrégats**	**Sur l'orientation de production**	**Sur la typologie**	**Sur l'orientation de production**	**Sur la typologie**
1-3-5-2-2-1	148	87,8%	44,6%	94,6%	62,8%
1-0-1-1-0-5	195	87,4%	50,3%	84,9%	50,3%

Tableau 5 : Résultats comparatifs de deux méthodes sur l'ensemble du Tessin (détail en annexe VI.4-5)

Dans toutes les phases de ce travail, nous évaluons la qualité de l'agrégation sur plusieurs niveaux : tout d'abord, nous observons le nombre total d'agrégations, qui doit avoisiner les 900 à 1000 opérations pour toute la Suisse. Ensuite, nous observons le taux de changement induit par l'agrégation sur l'expression de la variable à cartographier. Nous mettons en évidence, pour les communes agrégeantes et agrégées, celles qui ne changent pas de typologie, et celles qui conservent une orientation de production cohérente (si la typologie ne change pas de classe, végétale, animale ou mixte, selon le tableau 1).

7 SIGNIFICATION DES INDICATEURS

Suite à la programmation des différents indicateurs et à l'exécution du programme multicritère ainsi déterminé, il s'est posé la question de l'interdépendance des indicateurs choisis. Comme nous cherchons un jeu de paramètres porteurs d'information nécessitant un minimum d'opérations informatiques pour un résultat satisfaisant, nous nous sommes alors attardés sur la définition des indicateurs traduisant concrètement et mathématiquement les objectifs dessinés par les critères.

Nous posons tout d'abord la question de l'importance du critère d'appartenance au canton, puisqu'elle est mise en doute déjà lors de l'établissement initial des indicateurs. Ensuite, nous cherchons à vérifier une idée survenue rapidement : le nombre d'exploitations comme indicateur de ressemblance est-il superflu ? Finalement, nous nous sommes penchés sur les indicateurs de forme, pour attester ou infirmer une redondance entre l'indice de compacité de Gravélius et la longueur de la frontière commune.

7.1 APPARTENANCE AU CANTON

L'expérience des premières opérations sur les communes nous montre que le critère d'appartenance semble induire plus de problèmes qu'il n'en résout. En effet, le traitement des communes enclavées devient tout simplement irréalisable, ayant pour conséquence le non respect des conditions sur la protection des données. De plus, ces cas particuliers représentent rarement des valeurs statistiques capables d'influencer significativement les résultats d'analyse statistique au niveau cantonal. Une liste des communes se trouvant isolées dans un autre canton est présentée en annexe VII.1 et permet de se convaincre que l'intégration de leurs données à l'information globale d'un autre canton ne remet pas en cause la portée des chiffres cantonaux. Des discussions supplémentaires avec l'OFS Agr. nous confortent dans l'idée d'abandonner ce paramètre. Il s'avère en effet que l'agrégation de communes de cantons différents devient inévitable lorsque la totalité ou le fragment principal du territoire communal se situe à l'extérieur des limites globales du canton d'origine. Nous optons pour une solution intermédiaire en conservant cet indicateur, mais en lui attribuant un poids faible lors de l'exécution de la méthode multicritère.

En outre, nous pourrons éviter partiellement ce problème dans les applications à venir, puisque nous disposons du contour des communes (état 2000) utilisé par l'OFS. M. Steffen, qui s'occupe de la cartographie à l'OFS, nous fournit une représentation de la Suisse qui abolit quasi-totalement les enclaves, carte correspondant au niveau de généralisation N°4 établit par la section statistique de superficie de l'OFS. De plus, quelques retouches y sont encore apportées pour un dessin des frontières plus arrondi sur les rivières et plus rectiligne sur les crêtes montagneuses.

7.2 RESSEMBLANCE

Nous nous sommes demandé s'il n'y avait pas une répétition dans le fait d'utiliser le nombre d'exploitations pour sélectionner les germes d'une part et pour le critère de ressemblance d'autre part, d'autant plus que nous employons également ce dernier paramètre comme indicateur de taille. Il est vrai que la comparaison d'entités de même taille se révèle plus significative que lorsque la différence d'entreprises agricoles est importante, mais la résolution multicritère devrait en tenir compte. En effet, la combinaison des facteurs

"orientation de production" et "taille" devrait être suffisante, d'autant plus que la valeur limite pour le critère de confidentialité est modulable.
Nous proposons de comparer l'exécution de référence du programme multicritère avec une application sans cet indicateur de ressemblance correspondant respectivement aux poids (1-3-5-2-2-1) et (1-0-5-2-2-1). Pour chaque réalisation, nous avons successivement attribué les valeurs 2, 4, 8, 12 et 12 comme indicateur de sélection des germes.

Les résultats sis en annexe VIII.1 démontrent qu'il est utile de ne considérer que l'orientation de production comme indicateur de ressemblance. Avec ces chiffres, deux arguments supplémentaires favorisent l'éviction du nombre d'exploitations. D'une part, la quantité d'agrégats augmente légèrement (+43 à 833) tout en maintenant élevée (88 et 93%) la proportion de communes d'orientation inchangée. D'autre part, le poids ainsi accordé aux indicateurs d'esthétisme influence favorablement la forme des agrégats puisqu'on retrouve nettement moins de polygones dont l'indice de Gravélius est supérieur à 1.6., et un indice maximum de 2.92 au lieu de 3.28. Remarquons à ce stade que de telles valeurs ne devait pas être atteintes avec les contours des communes 2000 fournis par l'OFS.

Nous décidons d'exclure cet indicateur pour les applications à venir sur les données 2000.

7.3 ESTHÉTISME

La plus grande partie des tests effectués sur le programme de base s'attache à évaluer la redondance entre l'indice de Gravélius et la frontière commune. Conceptuellement, nous pensons que ces deux indicateurs sont concordants, c'est-à-dire qu'ils mènent à des solutions et des agrégats très semblables. En effet, il semble manifeste que plus la portion de frontière partagée entre deux polygones est élevée, moins la forme globale de l'agrégat risque d'être alambiquée.

De manière à vérifier ou rejeter cette hypothèse, nous exécutons le programme, avec l'orientation de production comme champ discriminant lorsque deux voisins présentent le même score, selon la séquence suivante :

- 1e itération avec la valeur limite du nombre d'exploitations pour le critère de confidentialité, VL = 2,
- 2e itération avec VL = 4,
- 3e itération avec VL = 8,
- 4e itération avec VL = 12,
- 5e itération avec VL = 12, afin qu'il ne reste qu'une vingtaine de germes.

1. Test A, pour évaluer l'impact de chaque indicateur sur le résultat final par rapport à un essai sans ces indicateurs. D'autre part, cette expérience permet d'estimer l'influence de la combinaison des deux indicateurs par rapport à une situation où un seul indicateur de forme prend pratiquement le même poids que la combinaison.
    a. poids 1 3 5 2 2 1 jeu de référence
    b. poids 1 3 5 3 0 1 accent sur la frontière commune
    c. poids 1 3 5 0 3 1 accent sur l'indice de Gravélius
    d. poids 1 3 5 0 0 1 sans indicateur de forme

2. Test B : pour vérifier dans quelle mesure l'indice G et la longueur de la frontière commune (LFC) sont corrélés.
    a. poids 0 0 0 1 1 0
    b. poids 0 0 0 1 0 0
    c. poids 0 0 0 0 1 0

Pour chaque test, nous observons les résultats suivants :
- l'indice de Gravélius minimum atteint lors de l'agrégation (Min)
- l'indice maximum (Max)
- la proportion d'agrégats présentant un indice G < 1.2
- la proportion d'agrégats présentant un indice G > 1.6
- la proportion de communes agrégées dont l'orientation de production (cult. permanentes, végétale, animale, mixte) n'a pas changé (F. inch.)
- la proportion de communes agrégeantes dont l'orientation de production est conservée (To inch.).

Comme le montrent les résultats numériques du test A présentés en annexe VIII.2, un critère de forme est tout de même nécessaire. De plus, la combinaison des indicateurs constitue un compromis optimisé entre les résultats obtenus par les solutions individuelles, même si la contribution à l'amélioration du résultat n'est pas manifeste.
En comparant les applications du test A sans indicateurs de forme avec celles où un seul indicateur est utilisé, il n'est pas évident de déceler un effet prépondérant de l'indice G ou de la frontière commune. Certes l'amélioration est sensible avec l'indice G sur l'ensemble des agrégats et perceptible avec la LFC sur les formes déjà assez régulières, mais on ne peut pas vraiment affirmer qu'un indicateur soit plus adapté. Il semble tout de même que l'indice G soit plus puissant tout en ne perturbant qu'un minimum l'action du critère de ressemblance comme on peut le constater sur le nombre de communes dont la classe de production reste inchangée.

Le test B (annexe VIII.3) démontre une puissance supérieure de la part de l'indice de Gravélius : utilisé comme seul facteur d'agrégation, il permet de réduire de 2/3 le nombre d'agrégats ayant un indice G > 1.6 par rapport à l'application "normale". Cependant, le double de communes changent de typologie durant ce processus. Même utilisée seule, la LFC ne parvient guère à construire des agrégats de forme régulière, si l'on compare avec l'application du programme sans indicateur de forme. Ce test confirme également que la combinaison des indicateurs représente une sorte de compromis sur les solutions particulières.

Compte tenu de ces observations, nous décidons d'éliminer un indicateur de forme, afin de ne pas rallonger inutilement les temps de travail. L'avantage de l'indice G de Gravélius, c'est sa robustesse face à des situations défavorables. En effet, par sa formulation mathématique, l'indice G est plus contraignant alors que l'indicateur "frontière commune" risque plus de laisser se former des agrégats indésirables. Sinon, les temps de travail doivent être équivalents, et les deux paramètres possèdent l'avantage d'être indépendants des autres indicateurs.
Pour les applications à venir sur les données du recensement 2000, nous n'utilisons plus que l'indice de Gravélius comme indicateur de forme.

L'annexe VIII.4 propose quelques exemples tirés des résultats graphiques de l'exécution du programme selon les circonstances, et illustrant assez bien les relations entre les indicateurs de forme.

Au vu des expériences et résultats présentés ci-dessus, nous décidons que pour l'ensemble des opérations réalisées sur la base des données 2000, il devient nécessaire

- de réduire fortement l'importance du critère d'appartenance au canton,
- d'éliminer la composante "Nombre d'exploitations" pour le critère de ressemblance,
- de supprimer l'indicateur sur la frontière commune pour la forme des agrégats et de ne conserver que l'indice de compacité de Gravélius.

En résumé, nous procédons, pour les données 2000 à l'élaboration d'une méthode d'agrégation sur la base des critères et indicateurs présentés dans le tableau ci-dessous :

Critères	**Indicateurs**
Confidentialité	Nombre d'exploitations agricoles
Ressemblance	Typologie de l'orientation de production
Appartenance au canton	Canton souverain
Esthétisme	Forme de l'agrégat (Indice de Gravélius) Taille de l'agrégat

Tableau 6 : Critères et indicateurs retenus pour les applications sur les données 2000

A noter que l'indicateur de confidentialité n'intervient pas directement dans le processus d'agrégation puisqu'il sert à sélectionner les germes, ainsi qu'à assurer la pertinence des agrégats.

Avant d'entamer la manipulation des données 2000, nous cherchons à développer d'autres méthodes d'agrégation basées sur des indicateurs statistiques. Quelques premiers tests sont comparés aux résultats de référence obtenus par le programme de résolution multicritère.

8 AUTRES PROCÉDURES D'AGRÉGATION

8.1 MÉTHODES STATISTIQUES

Une première sélection des méthodes statistiques applicables dans le cadre de ce projet est décrite dans le chapitre 2. Nous avons alors retenu deux types de corrélation, mis en œuvre sur les observations et sur les rangs respectivement. Nous proposons de réaliser quelques expériences pour déterminer la qualité d'agrégation de ces méthodes. Nous comparons les résultats de ces processus avec ceux de la résolution multicritère. Pour ces systèmes basés sur la corrélation, le principe d'agrégation est simple : le voisin qui obtient le coefficient le plus élevé devient la commune agrégeante qui annexe le germe et ses attributs.

8.1.1 CORRELATION SIMPLE SUR LES OBSERVATIONS

En première itération, avec une valeur limite du nombre d'exploitations fixé directement à 12, nous obtenons les résultats suivants :

Seuil de Corrélation	Nombre d'agrégations
r = 0.9	36
r = 0.8	82
r = 0.7	111
r = 0.6	147

Tableau 7 : Corrélation simple : résultats d'une itération, avec seuil de corrélation variable

A titre de comparaison, dans les mêmes conditions, la méthode multicritère effectue 479 agrégations.

Dans une autre approche, nous exécutons un programme avec le coefficient de corrélation comme unique facteur d'agrégation. Le seuil pour le coefficient de corrélation est fixé à 0.8, pour 5 itérations où la valeur limite du nombre d'exploitations prend les valeurs 2, 4, 8, 12 et 12 successivement.

Statistiques	Corrélation	Multicritère
Nombre total d'agrégats	125	795
Communes agrégées inchangées	44.8 %	86.9 %
Communes agrégeantes inchangées	82.4 %	92.2 %

Tableau 8 : Corrélation simple : résultats comparatifs avec la méthode multicritère sur toute la Suisse

Au vu des résultats présentés dans le tableau ci-dessus, il semble évident que la méthode utilisant la corrélation sur les observations ne peut être utilisée comme seul facteur d'agrégation sur les données des communes. Le nombre de variables est probablement adéquat, mais les échantillons représentant les germes ne sont pas suffisamment caractéristiques ou typés : il est très ardu d'établir un indice de ressemblance avec une commune qui ne contient que 1, 2 ou même 5 exploitations. En effet, de telles communes sont décrites par un vecteur dont la majorité des 8 composantes sont nulles. Ainsi, de petites variations de structure du vecteur peuvent induire de forts changements du coefficient de corrélation.

8.1.2 CORRELATION SUR LES RANGS

L'avantage de manipuler les rangs par rapport aux observations est le fait de s'affranchir des valeurs nulles qui faussent passablement les calculs. Il est de rigueur, lorsque nous avons affaire à des valeurs identiques, de leur attribuer un rang moyen. Pourtant, nous n'allouons un rang moyen que pour les variables nulles, ce qui permet de diminuer suffisamment le hasard dans l'ordre d'attribution de ces rangs. Nous pouvons alors espérer que cette méthode possède un pouvoir discriminant plus marqué. Pour s'en rendre compte, nous effectuons des tests identiques à ceux développés ci-dessus. Voici les résultats :

Seuil de corrélation	**Nombre d'agrégations**
$(1-\alpha) = 0.99 \rightarrow r = 0.88$	273
$(1-\alpha) = 0.98 \rightarrow r = 0.83$	348
$(1-\alpha) = 0.95 \rightarrow r = 0.74$	434
$(1-\alpha) = 0.90 \rightarrow r = 0.64$	497
$(1-\alpha) = 0.80 \rightarrow r = 0.52$	519

Tableau 9 : Corrélation sur les rangs : résultats d'une itération, avec seuil de corrélation variable

Nous pouvons déjà remarquer que le nombre d'agrégations est comparable à celui obtenu lors de la résolution multicritère (479), ce qui est favorable. La réalisation du second test, avec un degré de confiance de 0.95, soit un seuil de corrélation légèrement inférieur à la réalisation précédente, nous fournit les résultats suivants :

Statistiques	**Corrélation sur rangs**	**Multicritère**
Nombre total d'agrégats	664	795
Communes agrégées inchangées	83.4 %	86.9 %
Communes agrégeantes inchangées	88.9 %	92.2 %

Tableau 10 : Corrélation sur les rangs : résultats comparatifs avec la méthode multicritère sur la Suisse

Nous pouvons noter une bonne similarité entre les deux méthodes, ouvrant alors de nouvelles perspectives pour les traitements à venir, soit sur le type de production des entreprises agricoles en 2000, soit sur d'autres statistiques agricoles, économiques, démographiques, etc. Nous développons tout de même cette procédure sur la base des données 00.

8.2 CALCUL D'UNE DISTANCE

8.2.1 SCORE DE PEARSON (KHI-CARRE)

Le score de Pearson correspond à une mesure de distance entre deux polygones représentés par leur vecteur caractéristique. Dès lors, la sélection détermine comme commune agrégeante, celle qui présente la distance la plus faible par rapport au germe. Pour éviter les ambiguïtés dues à la présence de nombreuses valeurs nulles, nous écartons les communes adjacentes présentant également une distance nulle. Cette contrainte s'applique au détriment des voisins identiques, situation quasiment inexistante a priori.

Comme pour les méthodes statistiques, nous effectuons un premier test consistant en une itération, avec la valeur limite du nombre d'exploitations VL fixée à 12, expérience pour laquelle nous obtenons les résultats suivants :

Nombre d'agrégats effectués (Score > 0)	191
Nombre de germes écartés (Score = 0)	691

Tableau 11: Score de Pearson : résultats d'une itération

Dans les mêmes conditions (1 itération, VL =12), l'application multicritère de référence réalise 479 agrégats.
Au vu de ces quelques chiffres et de l'expérience réalisée sur la corrélation simple, nous ne nous hasardons pas à prolonger les tests sur cette méthode. Nous décidons d'abandonner cette voie, de la même façon que le calcul d'une distance sur la typologie, non retenu à cause de l'échelle discontinue et des différences d'intervalles pour chaque orientation de production.

8.2.2 DISTANCE SUR LES COMPOSANTES PRINCIPALES

Nous avons importé dans S-Plus le fichier contenant, pour chaque commune, le nombre d'entreprises agricoles par spécialisation, soit les 8 variables servant au calcul de la typologie. Nous avons réalisé une analyse en composantes principales, dont le premier résultat se compose des valeurs propres de la matrice de corrélation et de leur proportion par rapport à la somme de ces valeurs. Ce pourcentage nous renseigne sur la part d'information contenue dans chaque composante du nouveau jeu de variables. Nous obtenons les résultats suivants, dont le détail se trouve en annexe X.1 :

Composante principale	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Valeur propre	1.4887	1.306	1.072	0.938	0.802	0.732	0.705	0.614
Taux d'information	27.7 %	21.3 %	14.4 %	11.0 %	8.0 %	6.7 %	6.2 %	4.7 %
Taux cumulé	27.7 %	49.0 %	63.4 %	74.4 %	82.4 %	89.1 %	95.3 %	100 %

Tableau 12 : ACP : résultats pour les variables servant au calcul de la typologie

Comme mentionné au chapitre 2, il faut que la première valeur propre contienne au moins 60% de l'information donnée par les variables initiales, ou 85% pour les trois premières composantes. Malheureusement, cette expérience ne remplit aucune de ces deux conditions. Il est dès lors inutile de poursuivre les tests sur ce nouveau jeu de paramètres. Précisons que ce résultat était plus ou moins prévisible, puisque, intuitivement, les catégories d'orientation de production étaient déjà bien distinctes les unes des autres. Néanmoins, nous reprenons la même démarche avec les données 2000 de MBS pour chaque exploitation, même si la probabilité est assez forte pour que les conclusions soient identiques.

Pour les applications sur les données 2000, nous décidons de persévérer d'abord dans la voie multicritère pour les raisons suivantes :

a. Cette méthode est très proche de l'actuelle procédure manuelle ce qui rend sa compréhension très facile et une adaptation rapide à cette technologie
b. Cette technique est plus fine que la corrélation, puisqu'elle permet à l'opérateur de garder un contrôle sur une plus grande palette de paramètres d'agrégation
c. A développement égal, elle permet d'agréger plus de communes avec d'excellents résultats.

Malgré tout, l'adaptation du processus d'agrégation par corrélation sur les rangs semble ouverte à tous les domaines de la statistique avec un nombre de modification moindre, ce qui permettra probablement d'effectuer quelques tests sur les données du recensement 2000 des structures agricoles. De même, la détermination d'une distance entre le germe et chacun de ses voisins permettrait a priori d'élargir le nombre de méthodes capables de fournir des résultats satisfaisants.

9 DÉVELOPPEMENT DU PROTOTYPE SUR LES DONNÉES 2000

Comme mentionné au chapitre précédent, nous développons tout d'abord la méthode multicritère sur la base du jeu d'indicateurs réduit présenté ci-dessous.

Critères	**Indicateurs**
Confidentialité	Nombre d'exploitations agricoles
Ressemblance	Orientation de production (Marges brutes standard)
Appartenance	Canton souverain
Esthétisme	Forme de l'agrégat (Indice de Gravélius) Taille de l'agrégat (nombre d'exploitations)

Tableau 13 : Données 2000 : Récapitulatif des critères et indicateurs

Ensuite, nous explorons principalement deux voies pour la détermination du voisin le plus ressemblant. Nous poursuivons les recherches sur la méthode de la corrélation sur les rangs, et tentons de déterminer un indice de ressemblance basé sur un calcul de distance virtuelle (distance euclidienne…).

Pour l'état 2000, nous disposons du contour des communes généralisées (limites simplifiées) et, pour chaque exploitation, du numéro de commune, des marges brutes standard MBS primaires (P1 à P5) et secondaires (P11 à P131), des surfaces importantes du domaine agricole, du nombre de têtes de bétail et de la spécialisation de l'entreprise. Comme nous n'avons actuellement pas de rattachement spatial de l'exploitation, nous devons procéder à l'agrégation de ces données individuelles par commune, ce qui réduit l'éventail des applications possibles à partir d'informations aussi riches. Malgré tout, le fait de manipuler des données monétaires (MBS) présente l'avantage de la continuité de l'échelle, par opposition aux différents intervalles induits par l'établissement d'une typologie des orientations de production.

9.1 MÉTHODE MULTICRITÈRE

Pour cette procédure, différentes stratégies peuvent être envisagées. Tout d'abord, il est possible d'appliquer la même méthode que lors de la manipulation des données de 1996. En outre, à partir des MBS compilées, nous pouvons calculer la spécification de la commune et l'utiliser comme indicateur de ressemblance. Finalement, il s'avère certainement judicieux d'intégrer les procédures de corrélation ou de calcul de distances dans le programme multicritère, pas comme unique facteur d'agrégation, mais comme remplacement de l'indicateur de ressemblance. Cette dernière variante fait l'objet du chapitre 10.

9.1.1 TYPOLOGIE

Cette première expérience ignore l'information fournie par les MBS. Elle sert principalement à établir un jeu de poids adéquat qui pourra servir de référence pour toutes les manipulations ultérieures.

La phase initiale consiste, pour chaque commune, à sommer les exploitations selon chaque spécialisation, puis à insérer le résultat dans la base de données ArcView contenant les attributs des communes 2000, de manière à obtenir une sorte de mise à jour de la table issue du recensement 96 et utilisée dans la partie précédente de ce travail.

Nous expérimentons plusieurs jeux de poids dont nous comparons le résultat d'agrégation à la représentation des données originales, afin de déceler quelle solution se rapproche le plus de la réalité. L'annexe XII.1 (tableau 1) contient le schéma d'attribution des scores. Le tableau suivant présente les résultats chiffrés, alors que les annexes XIV.1-2-5 contiennent les détails correspondants et la carte finale de référence. L'ordre des poids représente respectivement les indicateurs "canton", "orientation de production", "forme" et "taille" de l'agrégat. D'autre part, nous exécutons le programme sur l'entier de la Suisse, en plusieurs itérations où seule la valeur limite du nombre d'exploitations varie (VL = 2, 4, 8, 12 et 12).

	Poids	1-1-1-1	1-3-1-3	1-5-3-1	2-3-1-1
	Nombre d'agrégats	863	837	903	870
Communes agrégées inchangées	Sur l'orientation de production	85,2%	86,9%	89,5%	88,7%
	Sur la typologie	50,2%	51,0%	52,2%	53,0%
Communes agrégeantes inchangées	Sur l'orientation de production	89,6%	90,7%	93,8%	93,0%
	Sur la typologie	65,2%	66,3%	71,0%	69,5%

Tableau 14 : Méthode multicritère sur la typologie : variations des poids et détermination du jeu de référence

Manifestement, le fait de récupérer la structure de référence employée sur des données 96 semble porter ses fruits, puisque c'est avec le jeu de poids 1-5-3-1 que nous obtenons les meilleurs résultats. De plus, ces chiffres sont même légèrement plus élevés que les valeurs obtenues lors des manipulations sur les données 96, tendant ainsi à démontrer que l'élimination de 2 indicateurs sur 6 était nécessaire, ou tout au moins profitable. En outre, si l'on observe également les statistiques sur la forme des agrégats (annexe XIV.1), il est manifeste que l'utilisation des limites généralisées des communes suisses améliore considérablement l'esthétisme et la clarté de la carte.

9.1.2 SPECIALISATION

Comme les diverses techniques suivantes, cette solution propose un test de ressemblance plus fin par rapport à la variante précédente, puisqu'elle s'appuie partiellement sur les données supplémentaires. Toutefois, la possibilité la plus simple de comparer les résultats des différentes méthodes développées consiste à maintenir dans la table des communes les champs nécessaires à la détermination de la typologie. De ce fait, cette table devient rapidement volumineuse, entraînant un accroissement du temps de calcul.

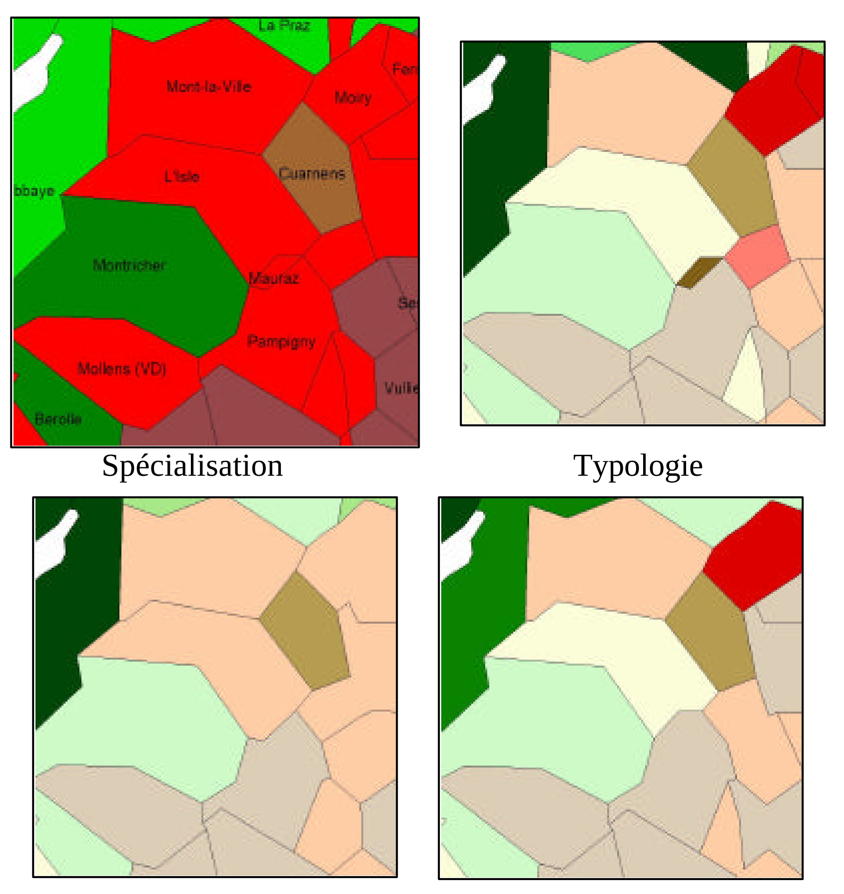
Dans ce cas précis, la table des communes contient en plus les MBS primaires et la spécialisation de l'entité communale. L'annexe XII.1 (tableau 2) contient le schéma d'attribution des scores pour l'unique test se déroulant dans les conditions standard : poids de référence (1-5-3-1) et VL variable (2, 4, 8, 12 et 12). La figure ci-après résume les résultats de cette opération, dont le détail se trouve en annexe XIV.3-4(carte).

		Communes agrégées inchangées		Communes agrégeantes inchangées	
Poids	Nombre d'agrégats	Sur l'orientation de production	Sur la typologie	Sur l'orientation de production	Sur la typologie
1-5-3-1	903	81,4%	48,3%	85,7%	63,9%

Tableau 15 : Méthode multicritère sur la spécialisation

On remarque une légère baisse générale de l'efficacité du programme par rapport à l'application de référence. Pourtant, il faut relativiser quelque peu cette première impression. En effet, la méthode de référence compare en quelque sorte la répartition communale des types d'entreprises. En agrégeant toutes les MBS des exploitations au niveau de la commune, on décrit de façon plus rigoureuse la structure de la production agricole de la commue. Il peut dès lors arriver que des communes de structure semblable (exprimée par le chiffre de la spécialisation) soient agrégées par ce second système, mais pas par la résolution de référence car la majorité des exploitations ne correspondent pas à cette spécialisation globale.

Dans l'exemple ci-dessous illustrant une telle situation, Mauraz est agrégée à Pampigny en suivant la typologie, à L'Isle en comparant la spécialisation.



Spécialisation Typologie

Les deux premières images présentent la situation initiale, du point de vue de la spécialisation et de la typologie respectivement.

L'Isle
spécialisation : mixte
typologie : majorité animale

Mauraz
spécialisation : mixte
typologie : végétale + 65%

Pampigny
spécialisation : mixte
typologie : majorité végétale

Vis-à-vis de ces images, en dessous, il y a les illustrations du résultat de l'agrégation selon chacune des méthodes, et présentées par la typologie de l'agrégat.

Figure 15 : Conflits entre agrégation basée sur la spécialisation et présentation selon la typologie des communes

Dans cet exemple, l'agrégat Mauraz – L'Isle prend une typologie mixte conformément à la spécialisation et à la structure de production de chacune des communes. Cependant, la statistique finale exprimera le changement de typologie des deux communes, tandis que, pour l'agrégat Muraz – Pampigny, seul le changement de typologie de Muraz sera exprimé. De telles situations sont cependant assez rares, n'altérant que faiblement la signification des

résultats chiffrés, dont l'ordre de grandeur est donc valable. Remarquons tout de même que 78% des communes agrégées et 91% des communes agrégeantes ont conservé une typologie proche de l'initiale. Cette différence dans le taux de changement d'affectation des communes nous incite à penser qu'il serait probablement profitable d'élaborer une méthode de classification intermédiaire entre la typologie très rigide, et la spécialisation assez générale et simplificatrice.

9.2 MÉTHODE DE LA CORRÉLATION SUR LES RANGS

Comme pour les données du recensement 96, nous appliquons cette technique comme unique facteur d'agrégation, à la différence près que, pour chaque commune, l'échantillon comporte 12 montants de MBS au lieu de 8 valeurs représentant le nombre d'exploitations par classe de production. Cette variante permet de mieux considérer la réelle structure agricole de la commune, indépendamment des particularités de chaque exploitation.

La décision d'agrégation dépend de deux paramètres : comme pour toutes les autres méthodes, un polygone voisin d'un germe ne sert de commune agrégeante que s'il est le plus ressemblant, c.-à-d. le polygone présentant le coefficient de corrélation r le plus élevé. De plus, afin d'améliorer la qualité des agrégats du point de vue de la ressemblance, nous appliquons une valeur limite du coefficient r au-dessous de laquelle la réunion des polygones n'a pas lieu. Nous réalisons diverses opérations à partir de ce programme, en variant le seuil de corrélation (r = 0.52, 0.74 et 0.88), lors de la procédure habituelle d'itérations avec la valeur limite du nombre d'exploitations mobile (VL = 2, 4, 8, 12 et 12). Ensuite, nous expérimentons une formule différente, en fixant VL = 12 sur 5 itérations avec un coefficient de corrélation r = 0.74. Dans le tableau ci-dessous se trouve un résumé des résultats obtenus pour ces premières manipulations.

	Coefficient de corrélation r	**0,52**	**0,74**	**0,88**	**0,74 et VL = 12**
	Nombre d'agrégats	911	806	620	835
Communes agrégées inchangées	**Sur l'orientation de production**	78,7%	79,3%	80,2%	80,4%
	Sur la typologie	43,8%	43,6%	43,6%	44,1%
Communes agrégeantes inchangées	**Sur l'orientation de production**	90,0%	91,4%	91,0%	90,1%
	Sur la typologie	67,5%	69,0%	66,5%	67,0%

Tableau 16 : Corrélation de Spearman : variations du seuil de corrélation : exemple 1

Etonnamment, toutes ces formules aboutissent à des résultats très semblables, alors qu'on pourrait s'attendre à un taux de réussite plus important pour un seuil élevé. On peut envisager une partie d'explication à cette situation : dans les Alpes, les Préalpes et la partie supérieure du Jura, l'élevage est presque exclusif; les coteaux du Valais et du Tessin, ainsi que des bords du lac Léman sont voués à la viticulture; la plaine s'appuie sur les grandes cultures ou sur des exploitations mixtes. Cette sorte de régionalisation de la production diminue le risque de conflits d'agrégation, expliquant le bon taux général de réussite, même avec un seuil de corrélation bas (voire nul comme ci-dessous). D'autre part, la frontière entre les typologies peu marquées (mixte, majorité animale ou végétale) est relativement mince, permettant à des communes de typologie différente d'obtenir un coefficient de corrélation assez élevé. On se retrouve dans une situation très semblable à celle présentée au paragraphe 9.1.2 et illustré par la figure 15.

Il faut encore préciser que, après la première opération (r = 0.52), il reste 38 germes dont tous les voisins présentent un coefficient inférieur au seuil fixé, 160 après la seconde et la quatrième, 383 après la troisième. Comme il reste passablement de germes ne pouvant pas être agrégés au vu des conditions imposées par le seuil de corrélation, nous tentons deux alternatives. D'une part, nous décidons d'exécuter le programme sans seuil (r = 0) avec VL croissante, et d'autre part nous reprenons les résultats de la deuxième expérience pour y appliquer le même régime (r = 0) sur deux itérations supplémentaires avec VL = 12.

	Coefficient de corrélation r	0	0,74 puis 0
	Nombre d'agrégats	943	960
Communes agrégées inchangées	Sur l'orientation de production	78,5%	77,3%
	Sur la typologie	43,1%	42,3%
Communes agrégeantes inchangées	Sur l'orientation de production	89,4%	91,0%
	Sur la typologie	66,6%	69,9%

Tableau 17 : Corrélation de Spearman : variations du seuil de corrélation : exemple 2

Au vu de ces résultats (détails en annexes XII.2 et XV.1-2), même s'ils sont légèrement inférieurs à la solution multicritère de référence, nous pouvons affirmer que nous tenons là une variante assez efficace ouvrant de grandes perspectives de généralisation. En effet, nous pouvons aisément imaginer développer cette méthode pour la rendre polyvalente, c.-à-d. applicable avec d'autres jeux de données, issues de la statistique agricole ou non.

9.3 MÉTHODE PAR CALCUL DE DISTANCE

Pour le calcul de distance, chaque commune est considérée comme un vecteur dont les composantes représentent les cinq MBS primaires. Nous sommes contraints de normaliser les données transmises par l'OFS, pour que, lors de l'agrégation, il n'apparaisse aucun biais dû à la taille respective des exploitations. Nous avons alors l'assurance que la signification est identique pour des distances mesurées entre communes de grandeurs semblables ou très différentes. Nous testons deux manières de normaliser ces données : d'une part, nous réduisons chaque MBS à son pourcentage par rapport au total de ces gains potentiels, ce qui correspond à la méthode la plus fréquente. D'autre part, nous transformons le jeu de variables initiales (ici les MBS) en composantes principales issues du traitement S-Plus visant à obtenir des paramètres indépendants.

La distance peut être définie de plusieurs manières (W. N. Venables & B. D. Ripley, 1994) :

- Euclidienne : racine de la somme des carrés des composantes
- Maximum : valeur maximale des différences des composantes, en valeur absolue
- Manhattan : somme de la valeur absolue des différences des composantes
- Binaire : proportion de valeurs non nulles que deux vecteurs n'ont pas simultanément en commun, soit le nombre de binômes avec un zéro et une valeur non nulle, divisé par le nombre de couples avec au moins une valeur non nulle.

Dans le cadre de ce travail, nous nous contenterons de la formule la plus courante, à savoir la distance euclidienne. La méthode "Manhattan" semble quasiment identique, alors que la

variante "Binaire" nous paraît moins bien adaptée à la problématique. Un test pourrait être réalisé lors de développements futurs avec la méthode "Maximum".

9.3.1 DISTANCE SUR LES MBS EN POURCENT

De la même manière que pour la méthode de corrélation, nous utilisons l'indicateur de la distance comme unique facteur d'agrégation. L'expérience de la corrélation ayant illustré que la mise en place d'un seuil s'avère insignifiante, nous effectuons un seul test avec la formule suivante : sera commune agrégeante, le voisin qui présentera une distance euclidienne minimale par rapport au germe (VL = 2, 4, 8, 12, 12 et 12). Ensuite nous tentons d'affiner quelque peu la procédure d'agrégation en calculant la distance sur un nombre restreint de composantes du vecteur caractéristique, en fonction de la typologie du germe. La distance D est déterminée sur la base des trois premières composantes si le germe est de typologie végétale, sur les deux dernières s'il est plutôt orienté sur la production animale, sur les cinq sinon. On peut également appliquer un système similaire en se fondant sur la spécialisation du germe. Le fonctionnement de ces deux procédés est détaillé en annexe XVI.1.

	Type de procédure	Normale	Fonction de la typologie du germe	Fonction de la spécialisation du germe
	Nombre d'agrégats	947	939	958
Communes agrégées inchangées	Sur l'orientation de production	76,6%	57,8%	78,9%
	Sur la typologie	39,6%	38,1%	44,8%
Communes agrégeantes inchangées	Sur l'orientation de production	85,1%	86,4%	90,1%
	Sur la typologie	57,3%	60,6%	66,6%

Tableau 18 : Calcul de distance euclidienne selon différents schémas, fonctions de la typologie ou de la spécialisation du germe

Les remarques effectuées lors des variantes précédentes (spécialisation et corrélation) peuvent s'appliquer telles quelles à propos de ces résultats. En effet, ceux-ci présentent un ordre de grandeur significatif rendant tangible une puissance plus faible de ce système par rapport à la résolution multicritère de référence. Malgré tout, la distance euclidienne constitue un moyen formidable et très simple de confronter deux vecteurs caractéristiques des communes à comparer. Aussi est-il possible d'appliquer cette méthode dans des situations diverses moyennant un minimum de modifications.

9.3.2 DISTANCE SUR LES COMPOSANTES PRINCIPALES

Nous avons importé dans SPlus le fichier contenant, pour chaque exploitation, les données exprimant les cinq MBS primaires. Nous avons réalisé une analyse en composantes principales, dont le premier résultat se compose des valeurs propres de la matrice de corrélation et de leur proportion par rapport à la somme de ces valeurs. Ce pourcentage nous renseigne sur la part d'information contenue dans chaque composante du nouveau jeu de variables. Nous obtenons les résultats suivants (cf. annexe X.2):

Composante principale	1	2	3	4	5
Valeur propre	1.112	1.049	0.994	0.969	0.858
Taux d'information	24.7 %	22.0 %	19.7 %	18.8 %	14.8 %
Taux cumulé	24.7 %	46.7 %	66.4 %	85.2 %	100 %

Tableau 19 : ACP : résultats pour les marges brutes standard primaires sur les exploitations

Comme mentionné au chapitre 2, il faut que la première valeur propre contienne au moins 60% de l'information donnée par les variables initiales, ou 85% pour les trois premières composantes. Malheureusement, cette expérience ne remplit aucune de ces deux conditions. Il est dès lors inutile de poursuivre les tests sur ce nouveau jeu de paramètres, comme nous l'avons supposé au chapitre 8 déjà sur la première exécution de ce type d'analyse.

D'autre part, le texte complet des programmes d'agrégation multicritère simple pour la cartographie de l'orientation de production en 1996 et en 2000 est disponible dans un document à parallèle à ce rapport.

10 GÉNÉRALISATION

Même si ce projet se rattache principalement à l'OFS, nous estimons que la problématique de représentation cartographique sous une contrainte (ici la confidentialité) constitue un thème relativement fréquent. De même, l'agrégation de polygones sur la base de leurs attributs représente une thématique importante dans le domaine de la télédétection numérique (polygone = pixel). Ainsi, une notion présente en arrière plan dans l'ensemble de ce travail réside dans la proposition d'une démarche générale dont le principe pourrait s'appliquer à d'autres domaines. Ce n'est pas tant l'agrégation "physique" des polygones et de leurs attributs qui importe, mais plutôt la sélection, parmi les communes adjacentes, de celle qui présente les caractéristiques les plus proches de la commune concernée (germe). Nous cherchons une formule capable de classer les voisins dans l'ordre de ressemblance, sur la base de plusieurs facteurs, et ceci dans des configurations diverses. Dans ce sens, la méthode de résolution multicritère semble la plus universelle puisqu'elle permet, d'une part, de traiter plusieurs variables simultanément dans une procédure assez simple. D'autre part, ce système parvient à réunir, dans un jeu d'indicateurs comparables, des données de natures très diverses, qualitatives ou numériques, d'échelles et/ou d'unités différentes, etc. Très modulable donc, cette formule s'adapte à toutes les situations, mais nécessite à chaque fois la définition particulière des indicateurs et de leur importance au travers d'un jeu de poids. C'est pourquoi nous avons également testé des méthodes moins souples puisqu'elles nécessitent une uniformité d'expression des variables à analyser. En effet, tant la corrélation que le calcul d'une distance exigent que les composantes des vecteurs caractérisant les objets à comparer soient exprimées dans la même unité et sur des échelles semblables. Ce type de méthode présente l'avantage d'être répétitives : une fois la structure du programme définie, la procédure peut fonctionner indépendamment du nombre de paramètres ou de leur unité. A l'inverse de la variante multicritère où l'on tente d'élaborer un jeu d'indicateurs restreint, un nombre élevé de variables permet à ces méthodes (surtout la corrélation) d'avoir un caractère discriminant plus fort. De telles méthodes sont également applicables dans des situations diverses, même si elles ne considèrent qu'un facteur d'agrégation.

Ainsi, la solution consistant à regrouper les différentes méthodes testées pourrait probablement s'appliquer à d'autres domaines que la statistique agricole, moyennant les modifications inhérentes aux particularités de toute discipline.

10.1 COMBINAISON DE MÉTHODES

Pour vérifier ces propos, nous effectuons des expériences consistant à injecter la corrélation ou le calcul de distance dans le jeu d'indicateurs de la méthode multicritère, en tant qu'unique indicateur de ressemblance. Dans les deux cas, nous remplaçons l'attribution discrète d'un score par le système suivant : on alloue le score maximal au(x) voisin(s) qui obtient (obtiennent) la corrélation la plus élevée (ou la distance la plus faible), et, aux autres, un score dégressif dépendant du nombre de voisins restants. Cette formule permettra peut-être d'augmenter le pouvoir discriminant de la méthode, étant de surcroît appliquée également aux indicateurs d'esthétisme (taille et indice de Gravélius de l'agrégat). Pour la comparaison, nous effectuons également un test avec une attribution échelonnée des scores telle qu'elle a été appliquée jusqu'ici. Les schémas d'allocation des scores ainsi que le principe du score dégressif sont présentés en annexe XII.3. Remarquons encore que ce système réduit quelque peu le caractère empirique de l'attribution des scores. En effet, seules les valeurs maximum et minimum de l'échelle sont décidées par l'homme. Ensuite, le score n'est pas dépendant d'intervalles où situer la valeur de l'indicateur, mais correspond à l'ordre décroissant des

résultats de chaque commune sur cet indicateur. D'autre part, le fait d'utiliser un indice de ressemblance (corrélation ou distance) permet de s'affranchir du traitement au cas par cas et de décrire le degré de similitude sur la base des valeurs réelles caractérisant la communes.

Nous utilisons à nouveau les paramètres de référence pour cette application, à savoir les poids 1-5-3-1 et une valeur limite du nombre d'exploitations variable : VL = 2, 4, 8, 12 et 12. Le coefficient de corrélation se détermine, comme précédemment, selon la formule de Spearman. De même, nous calculons la distance euclidienne sur toutes les composantes du vecteur caractéristique des communes. Le tableau suivant résume les résultats obtenus détaillés en annexe XVI.2-3(carte).

	Type de procédure	Combinaison multicritère + corrélation		Combinaison multicritère + distance		Référence
	Score	Echelonné	Dégressif	Echelonné	Dégressif	Echelonné
	Nombre d'agrégats	917	943	878	923	903
Communes agrégées inchangées	Sur l'orientation de production	78,2%	77,8%	80,8%	79,5%	89,5%
	Sur la typologie	44,6%	43,2%	45,1%	45,9%	52,2%
Communes agrégeantes inchangées	Sur l'orientation de production	85,7%	86,6%	85,5%	88,6%	93,8%
	Sur la typologie	64,3%	64,3%	61,5%	67,9%	71,0%

Tableau 20 : Combinaison de méthodes : multicritère + distance ou corrélation

Comme nous pouvions nous y attendre, cette dernière modification ne révolutionne pas la méthode d'agrégation ni son résultat. Dans l'ensemble, les résultats sont bons, malgré une légère baisse de puissance par rapport à la solution de référence. La combinaison des procédures n'améliore pas l'efficacité de l'agrégation, mais retient les avantages de chaque composante, surtout lorsque le critère de ressemblance est caractérisé par de nombreuses variables. Dans de tels cas en effet, tant la corrélation que la distance évitent de régler une à une toutes les alternatives, et la partie multicritère du programme permet de considérer également d'autres facteurs (esthétiques, arbitraires, etc.). Si, au contraire, on ne dispose que de peu d'éléments descriptifs, on atteint les limites de la méthode de la corrélation et la détermination d'une distance s'apparente beaucoup à l'attribution discrète d'un score. C'est pourquoi nous proposons de conserver cette combinaison multicritère + distance pour un prochain test basé sur d'autres données de la statistique agricole.

Auparavant, pour démontrer que les méthodes présentées dans ce travail sont finalement assez proches et que la méthode retenue offre le meilleur compromis entre généralisation de la procédure et application multifactorielle, nous effectuons quelques comparaisons supplémentaires. Tout d'abord, nous nous intéressons à connaître la taille moyenne des agrégats selon deux composantes : en premier lieu, nous observons combien chaque agrégat regroupe de commune originales. D'autre part, nous cherchons à savoir combien de types de production distincts sont finalement absorbés par chaque agrégat. Le tableau suivant contient un résumé de l'annexe XVIII.1 qui présente le détail des résultats.

Méthode	Nombre moyen de communes par agrégat	Nombre moyen de types de production par agrégat
Typologie (multicritère)	1.463	1.237
Corrélation sur les rangs	1.504	1.298
Distance euclidienne	1.492	1.329
Spécialisation (multicritère)	1.464	1.266
Combinaison multicritère et corrélation	1.501	1.297
Combinaison multicritère et distance	1.478	1.273

Tableau 21 : Comparaison des méthodes selon le nombre moyen de communes et de types de production différents regroupés dans chaque agrégat

Ce tableau montre encore une fois que la méthode discrète (attribution multicritère des scores utilisant la typologie ou la spécialisation comme base de ressemblance) fournit les résultats d'agrégation modifiant le moins la signification de la représentation cartographique. Nous pouvons également remarquer que la combinaison de la résolution multicritère et du calcul de distance constitue une solution favorable pour tout type de données.
Cette dernière impression se trouve confirmée par le tableau suivant (cf. annexe XVIII.2) présentant l'indice de Gravélius et le nombre d'exploitations moyens des agrégats différents entre la méthode de référence et chacun des autres systèmes. Le nombre d'agrégats identiques donne une indication sur la similarité des variantes.

Méthodes	Nb agrégats identiques	Agrégats différents			Tous les agrégats		
		Nombre	Indice G	Taille	Nombre	Indice G	Taille
Corrélation	1270	656	1,48	27,8	1926	1,36	36,6
Typologie		710	1,37	25,7	1980	1,39	35,6
Distance	1135	806	1,49	27,2	1941	1,41	36,3
Typologie		845	1,37	26,0			
Spécialisation	1557	426	1,36	24,9	1983	1,35	35,6
Typologie		423	1,38	25,1			
Multicorrélation	1312	617	1,40	27,0	1929	1,37	36,6
Typologie		668	1,38	24,9			
Multidistance	1348	611	1,40	26,5	1959	1,37	36,0
Typologie		632	1,38	25,6			
Communes originales		Moyenne 1400	Moyenne 1,28	Moyenne 11,7	2896	1,31	24,3

Tableau 22 : Nombre d'agrégats, indice de Gravélius et nombre d'exploitations moyens par agrégat : sur toute la Suisse et sur les régions où les différentes méthodes n'ont pas effectué les mêmes agrégations.

Nous pouvons affirmer que la combinaison multicritère + distance représente l'aboutissement de ces premiers développements. Toutefois, dans l'optique d'une utilisation modérée d'un tel programme, les résultats obtenus par la simple résolution multicritère en comparant les communes sur leur classe de production plutôt que sur leur typologie laissent envisager quelques améliorations possibles. En effet, malgré la simplicité de la classification des communes d'après la méthode utilisée pour déterminer la spécialisation des exploitations, les produits de l'agrégation sont très semblables à la solution de référence. Nous pouvons dès lors imaginer qu'en élaborant un système de classification mixte entre la typologie et la spécialisation, nous puissions trouver une alternative aux répartitions appliquées actuellement. D'autre part, texte complet du programme d'agrégation par combinaison des méthodes multicritère et de calcul d'une distance euclidienne, pour la cartographie de l'orientation de production en 2000, est disponible dans un document parallèle à ce rapport.

10.2 APPLICATION DE L'AGRÉGATION DE COMMUNES SUR LA BASE D'AUTRES DONNÉES

Pour mettre en évidence le caractère modulable de la solution retenue (combinaison des méthodes multicritère et de calcul de distance), nous l'expérimentons sur un autre jeu de donnée, à savoir la représentation de la proportion communale de surfaces herbagères par rapport à la Surface Agricole Utile (SAU). Nous avons choisi ce thème en fonction des cartes illustant les "Reflets de l'agriculture suisse en 1996" (OFS, 1997), pour montrer l'utilité d'une telle méthode d'agrégation semi-automatique.

Dans cet exemple, la ressemblance n'est évaluée que sur un seul élément, le rapport Surf. Herbagères / SAU. La classification des communes ou agrégats s'effectue dans 6 catégories :

Pourcentage de surfaces herbagères	**> 80 %**	**> 65 %**	**> 50 %**	**> 35 %**	**> 30 %**	**= 30 %**
Catégories	1	2	3	4	5	6

Tableau 23 : Catégories pour la représentation des surfaces herbagères des communes suisses

Nous vérifions si cette dernière méthode est appropriée dans une telle situation en appliquant également, à titre comparatif, la méthode de référence. Pour chacune des procédures, nous utilisons le jeu de poids habituel (1-5-3-1) en faisant varier la valeur limite du nombre d'exploitations VL. Selon une convention respectée par l'OFS ainsi qu'au niveau international, la confidentialité est respectée lorsque 4 exploitations au moins par communes remplissent le critère sur lequel est jugée la ressemblance. Nous avons donc compilé les valeurs de surfaces herbagères en ne retenant par commune que les exploitations qui en déclarent. La SAU est toutefois considérée sur l'ensemble du territoire communal. Dans le tableau suivant se trouvent les résultats de cette opération (détails en annexe XII.4(scores) et XX.1-2-3).

	Méthode	**Référence**	**Multicritère + Distance**
	Score	**Echelonné**	**Dégressif**
	Nombre d'agrégats	224	225
Communes agrégées inchangées	**Différence de catégorie = 1**	100,0%	100,0%
	Même catégorie	75,0%	72,9%
Communes agrégeantes inchangées	**Différence de catégorie = 1**	98,7%	98,2%
	Même catégorie	91,5%	83,6%

Tableau 24 : résultats des deux méthodes retenues pour l'agrégation de communes pour la cartographie des surfaces herbagères en rapport à la SAU totale.

Ces premiers résultats se révèlent prometteurs malgré la simplicité de la variable à cartographier. En effet, même si la ressemblance n'est évaluée que sur un seul paramètre, les remarquables taux de conservation de la catégorie prouvent que la méthode combinée peut s'appliquer en de nombreuses situations. Ces propos confortent la position privilégiée de ce système dans l'optique d'un développement plus poussé du prototype, vers un cercle élargi de variables à cartographier, et donc vers un nombre d'utilisateurs potentiellement plus élevé.

11 SYNTHÈSE

11.1 LES AVANTAGES D'UN PROGRAMME D'AGRÉGATION D'UNITÉS TERRITORIALES

Nous pouvons ressortir deux avantages importants pour la section d'agriculture et de sylviculture de l'OFS. En premier lieu, un tel programme permet de rendre accessibles bon nombre de cartes qui impliquaient un délai de livraison presque rédhibitoire vu le temps qu'il fallait consacrer à leur élaboration par voie humaine. Moyennant quelques ajustages, un tel système d'agrégation semi-automatique permet de raccourcir considérablement le temps d'édition de la carte. De plus, cette méthode présente l'avantage de beaucoup réduire la part de subjectivité liée inévitablement à la nature humaine, ce qui rend cette procédure cohérente et régulière dans toute sa période de travail.
Du fait de sa rapidité d'exécution, ce système permet également d'élargir l'éventail de produits cartographiques de l'OFS, tout en conservant le strict anonymat des données. La principale innovation consiste à pouvoir créer des cartes à une échelle très proche de celle de la commune. Jusqu'alors, les quelques cartes publiées dans les "Reflets de l'agriculture suisse" (OFS, 1997) représentaient les informations regroupées au niveau du district ou du canton, ce qui convient parfaitement à la grandeur des images illustrant cet ouvrage. Cependant, on peut légitimement envisager insérer des cartes de format supérieur (jusqu'à la taille d'une page A4) sans altérer la lisibilité de cette publication. Désormais, il est possible de se rapprocher du niveau communal pour certains thèmes. C'est le cas pour la thématique de référence utilisée dans ce travail, l'orientation de production des entreprises agricoles. En utilisant le prototype créé lors de ce projet, nous avons considéré l'esthétisme de la carte en plus de la confidentialité inhérente à la problématique. Nous avons fixé à 12 le nombre minimum d'exploitations par agrégat pour réduire à environ 2000 le nombre de communes et pour garantir l'anonymat des données cartographiées selon la typologie établie à l'OFS Agr. En effet, même pour les types "Prédominance" de production animale ou végétale, 4 exploitations au minimum correspondent au type principal. Nous pouvons ramener cette valeur limite à 4 entreprises du secteur primaire si nous choisissons de représenter réellement la structure de la production agricole (spécialisation de la commune). Le thème cartographié est alors très proche de l'activité agricole menée sur la commune, mais nous perdons alors l'information concernant la classe de production de chaque exploitation. Par cet exemple, nous montrons que l'avantage peut-être le plus utile à l'OFS Agr est le caractère modulable et reproductible de la méthode : le principe reste valable pour toutes sortes d'agrégation et pour diverses données. L'application du prototype pour la représentation des surfaces herbagères rapportées à la SAU totale de la commune apporte la confirmation son efficacité et de sa maniabilité.

11.2 INCONVÉNIENTS ET LIMITATIONS

La plus grande contrainte se situe au niveau du critère de confidentialité puisqu'il détermine en grande partie les statistiques qui ne pourront jamais être représentées à l'échelle communale. Par exemple, il existe certaines productions animales qui restent relativement marginales si bien qu'il faudrait agréger beaucoup trop de communes pour atteindre le seuil de 4 exploitations. Le résultat ne serait plus représentatif de la localisation de ce type de production. Comme cette condition d'anonymat n'est pas amenée à changer à l'avenir, il faudrait plutôt s'affranchir des limites administratives et politiques: commune, district, canton éventuellement. Nous en discuterons quelques aspects dans le chapitre 12.

Le problème consistant à former des régions représentatives s'immisce à plusieurs degrés dans le cadre de ce projet. En raison de la présence de la contrainte de confidentialité, l'agrégation n'effectue pas la réunion des mêmes communes d'une statistique à l'autre, et entre deux jeux de poids différents. Les configurations finales n'étant pas identiques, il n'est dès lors pas possible de réaliser des comparaisons ou des recoupements entre ces cartes thématiques. Afin d'utiliser au mieux les fonctionnalités offertes par le SIT en complément au programme d'agrégation, il serait profitable de former des régions de référence, entre les niveaux communal et de district. Quelques pistes sont envisagées dans les "Perspectives" (chapitre 12). Quelques problèmes de comparaison entre les différentes méthodes testées apparaissent à cause de la typologie utilisée pour la cartographie. En effet, le calcul de distance et la corrélation ne pouvant être simplement appliqués à la typologie, nous avons choisi de manipuler les données de MBS pour évaluer la ressemblance entre communes. Cette réorientation induit quelques conflits de représentation thématique puisque l'agrégation et la présentation du résultat ne se font plus sur la même variable.

Au niveau de la méthode et du programme, les limitations ne proviennent pas du logiciel puisqu'il recèle une multitude de fonctions assez puissantes et qu'une nouvelle version (ArcGIS 8.1) propose des développements supérieurs. Toutefois, plus on envisage de manipuler des données détaillées, – jusqu'à l'échelle de l'exploitation par exemple – plus la programmation devient complexe, rallongeant par conséquent les temps de calcul de manière non négligeable. Ce phénomène se trouve accentué au vu du peu d'expérience initiale de la programmation en langage Avenue et du délai somme toute limité pour ce projet. D'autre part, les méthodes de calcul de ressemblance par corrélation sur les rangs ou par mesure de distances sont forcément plus générales et moins sélectives que le traitement au cas par cas. Cette différence provient du fait que la similarité est évaluée sur un indice englobant toutes les variables et indépendant de la structure du vecteur caractéristique. Ainsi, de par la formulation quadratique de ces indicateurs, des communes présentant des vecteurs différents peuvent être considérés égaux alors qu'il en existe un manifestement plus proche du germe dans sa structure. Dans la même veine, le critère de ressemblance sur le nombre d'exploitations – écarté pour les applications sur les données 2000 – se retrouve implicitement dans de tels indices, puisque les distances calculées seront forcément plus grandes si l'on compare des communes de taille différente ou de grandeur semblable. A l'extrême, nous pouvons concevoir un voisin, dont le vecteur est un multiple de celui du germe (structure identique), se voir recalé au profit d'une commune adjacente divergeant du germe sur plusieurs points, mais plus proche quant au nombre d'exploitations.

En rapport aux données du recensement, nous pouvons déplorer peut-être la représentativité relative de quelques informations. En effet, tout relevé d'ampleur se doit de fixer une date de référence. Dans le cas du recensement des structures agricoles, les exploitants décrivent leurs cultures (type, surface…), leur bétail (type, nombre…), etc. tels qu'ils sont le jour de référence. Pourtant, pour évaluer la réelle production d'une entreprise, agricole ou autre, il serait probablement judicieux de prendre en compte l'année écoulée jusqu'au jour déterminé. Dans la situation propre à ce projet, l'expression de l'orientation de production ne subirait vraisemblablement que de faibles modifications, mais nous disposerions d'une somme supérieure d'informations autorisant à cartographier plus de thèmes porteurs avec un degré de fiabilité élevé.

En outre, la forme même des données induit quelques problèmes de représentation cartographique. En effet, le recensement s'attache à décrire la structure et la production des exploitations agricoles. Par conséquent, nous disposons d'un jeu ponctuel d'informations liées au territoire, à l'opposé de la statistique de superficie qui représente la couverture du sol

effective à une période déterminée sans forcément connaître son importance dans la structure de l'exploitation. L'entité communale s'impose de suite comme le plus simple moyen de lier ces informations à une unité spatiale. Ce faisant, il est dès lors concevable de réaliser des agrégations basées en premier lieu sur la proximité des éléments surfaciques à cartographier. Dans le chapitre 12, nous développons quelques idées pour modifier cette méthode de "spatialisation" de données ponctuelles.

11.3 TABLEAU SYNTHÉTIQUE DES MÉTHODES TESTÉES

Le tableau suivant présente une synthèse des méthodes évaluées sur la base des données du recensement 2000, en résumant quelques avantages et inconvénients majeurs.

Méthode	Description	Avantages	Inconvénients
Multicritère simple	▪ Attribution de scores échelonnés selon différents critères d'importances diverses. ▪ Agrégation avec la commune présentant le score total le plus élevé	▪ Très proche de la méthode manuelle ▪ Très bons résultats (peu de changements d'orientation de production entre la carte initiale et la carte des agrégats	▪ Long à programmer, même avec un nombre moyen de paramètres ▪ 1 programme particulier pour chaque thème à cartographier (indicateurs, poids, scores)
Corrélation de Spearman	▪ 1 seul critère pour l'agrégation : le coefficient de corrélation le plus élevé ▪ Pas de scores	▪ Simple à comprendre, à utiliser et à programmer ▪ Bons résultats ▪ Bonnes possibilités de généralisation à d'autres thèmes	▪ Efficacité dépendante du nombre de variables ▪ Moins bonne sensibilité dans les situations délicates ▪ Mise à l'écart de l'esthétisme de la carte
Distance euclidienne	▪ 1 seul critère pour l'agrégation : la distance euclidienne la plus faible, calculée sur les composantes du vecteur caractéristique ▪ Pas de scores	Idem Corrélation de Spearman	Idem Corrélation de Spearman
Combinaison Multicritère + Corrélation	▪ Insertion du coefficient de corrélation dans la méthode multicritère comme critère de ressemblance ▪ Score dégressif	▪ Méthode assez proche de la résolution manuelle ▪ Rigueur dans l'attribution des scores ▪ Possibilités moyennes de généralisation	▪ Moins bonne sensibilité dans les situations délicates
Combinaison Multicritère + Distance	▪ Insertion de la distance euclidienne dans la méthode multicritère comme critère de ressemblance ▪ Score dégressif	▪ Idem Multicritère + corrélation ▪ Meilleurs résultats avec un nombre de paramètres restreints	▪ Idem Multicritère + corrélation

Tableau 25 : Tableau comparatif des méthodes testées

12 PERSPECTIVES

12.1 OPTIMISATION DU PROTOTYPE

Les résultats obtenus dans le cadre de ce travail avec les développements réalisés sur le programme sont encourageants. Pourtant, il ne s'agit justement que d'un prototype, avec toute la phase d'améliorations possibles que ce terme sous-entend. En effet, le langage de programmation Avenue n'étant pas maîtrisé à l'entame du projet, la procédure élaborée n'est certainement pas idéale puisqu'elle se base sur les fonctions les plus simples. Elle nécessite d'être épurée, pour améliorer la fluidité et la rapidité d'exécution, probablement à l'aide de fonctions prédéfinies ou d'autres disponibles sur le site Internet d'ESRI.

D'autre part, la recherche de la meilleure méthode n'étant pas aboutie, il est difficilement concevable de créer un véritable outil cartographique "transmissible ou diffusable". Dans l'état actuel du projet, nous pouvons envisager au maximum la création d'une extension ArcView interne à l'OFS, qui impliquerait l'élaboration d'une description et d'un mode d'emploi détaillés de la méthode. De surcroît, la version 3.2 de ArcView ne bénéficiera que de très peu d'améliorations à l'avenir, l'essentiel des développements étant engagés sur le nouveau concept ArcGIS. Ainsi, la possibilité la plus simple de créer un outil durable et évolutif consisterait probablement à transcrire le programme dans un langage plus universel comme Visual Basic.

Dans une perspective "idéaliste", ces quelques travaux pourraient aboutir à un logiciel ou une extension ArcView reconnue par ESRI qui offrirait la possibilité d'agréger non seulement des polygones, mais également des pixels…En outre, il serait envisageable de proposer différentes méthodes d'agrégation au sein d'un même instrument.

12.2 MARCHÉS POTENTIELS

Moyennant quelques modifications, le programme actuel peut être appliqué sur de nombreuses statistiques agricoles, à partir du moment où les variables composant le vecteur caractéristique sont exprimées sur des échelles comparables. En effet, il paraît illusoire de calculer un pourcentage d'élevage de chaque type d'animal si le nombre de têtes de bétail constitue la seule information disponible : comment comparer 23 vaches laitières avec 200 poules pondeuses si ce n'est en référence au gain qu'elles peuvent rapporter à l'exploitant ? Une fois cette condition respectée, nous pensons que ce prototype peut s'appliquer à l'ensemble des statistiques relevées par l'OFS.

Parmi quelques débouchés potentiels offerts aux produits qu'un tel programme d'agrégation permettrait de réaliser, nous pouvons imaginer d'une part renforcer la diffusion de telles cartes dans les organes décisionnels des administrations fédérales, cantonales ou régionales. Au niveau du district, il serait possible de diversifier les thèmes cartographiés, mais pas de proposer des résultats à l'échelle communale car il existe probablement des agrégats chevauchant les frontières préfectorales. Si l'on se concentre aux partenaires de la section d'agriculture de l'OFS, on peut facilement prétendre renforcer la collaboration avec d'autres offices fédéraux ou cantonaux (OFAG, OFEFP), des organismes spécialisés (USP, FAT), la presse agricole (Terre et Nature, AgriHebdo), etc. La dimension pédagogique et instructive de ces cartes pourrait également bénéficier d'un développement conséquent. Par ailleurs, il ne serait peut-être pas utopique de voir certaines branches touristiques bénéficier de quelques cartes pour leur promotion.

12.3 MODIFICATIONS DE LA CLASSIFICATION

La typologie établie par l'OFS Agr. exprime bien, pour chaque commune, la proportion d'exploitations agricoles spécialisées dans chaque domaine de production animale, végétale, mixte ou de cultures permanentes. Pourtant, nous pensons qu'il serait intéressant de représenter plus fidèlement la réelle structure de la production agricole sur le territoire communal. Nous avons donc élaboré une proposition pour la caractérisation de l'activité agricole communale (cf. annexe XXI.1). Il s'agit du même principe que le calcul de la spécialisation des exploitations, réaménagé pour déterminer un nombre plus grand de catégories. Pourtant, même si l'échelle de typologie est continue, il n'est toujours pas possible de calculer une distance directement sur la typologie. Les différentes classes n'étant pas réparties de manière homogène (trois pour la production végétale et deux pour les animaux), il semble définitivement impensable d'utiliser uniquement la typologie pour déterminer le degré de similitude entre communes.

12.4 AFFRANCHISSEMENT DES LIMITES COMMUNALES

Comme amorcée au chapitre précédent, la question de l'extension spatiale des données du recensement met en doute l'efficacité des frontières communales pour la représentation des statistiques agricoles. Les informations d'une entreprise agricole sont attribuées à la commune où se situe le centre d'exploitation, à savoir le bâtiment rural principal. Pourtant, il arrive souvent que des paysans s'occupent de terrains sis sur le territoire d'une autre commune : dans quelques cas extrêmes, plus de 70% de la surface cultivée se situe à l'extérieur de la commune d'origine. Ainsi, les données collectées sont légèrement biaisées puisqu'elles ne représentent pas la véritable production agricole de la commune. De plus, les différences de taille de communes, tant au niveau de la surface que du nombre d'exploitations induisent des difficultés de comparaison. Ainsi, nous jugeons qu'il serait utile de déterminer des régions de taille semblable possédant une homogénéité accrue qui représenterait également une amélioration par rapport aux données de certaines communes. Le principal problème réside dans la création d'un lien entre les données du recensement et leur emprise spatiale, puisque ces informations sont liées à l'exploitation sans véritable connexion avec les parcelles concernées. Nous proposons dans la suite quelques pistes et leur "applicabilité".

Tout d'abord, un lien ponctuel peut être créé en référençant les données relevées à l'emplacement d'un bâtiment déterminé, dont les coordonnées géographiques suisses sont déjà disponibles (géocodage). C'est le premier pas d'une méthode qui nécessiterait une phase de spatialisation de données ponctuelles. Dans ce domaine, nous pensons à des systèmes permettant de créer des polygones autour de points de référence (polygones de Thiessen) indépendants des attributs des points. D'autre part, il existe une extension ArcView qui propose plusieurs procédés d'interpolation permettant directement de représenter un thème donné comme sont souvent régionalisés les phénomènes météorologiques. En effet, il serait envisageable d'assimiler les données de MBS à une mesure d'intensité de production animale ou végétale. Pourtant, il persisterait un problème de représentation avec une telle méthode. En effet, nous pouvons aisément imaginer que l'essentiel des bâtiments agricoles se trouve concentré dans la partie construite des villages. Le résultat d'interpolation ne correspondrait pas à la distribution réelle des domaines exploités, et l'agrégation, même si elle s'applique directement aux entreprises agricoles, ne serait pas plus significative que pour les communes. Une telle solution basée sur le géocodage des immeubles, ne semble pas favorable, même si nous procédons d'abord à l'agrégation entre les points sur la base de l'information complète fournie par leurs attributs, avant d'accomplir la spatialisation proprement dite.

Une alternative, qui semble a priori assez lourde, consisterait à se baser au départ sur la position des parcelles de l'exploitation. Les agriculteurs appliquant les normes de la PI

(production intégrée), entre autres, doivent fournir les plans de leurs champs et la rotation des cultures mise en place. Actuellement, aucun référencement géographique suisse n'est disponible rapidement puisque les documents transmis par les agriculteurs sont des extraits – sur support papier – du cadastre communal, mais nous pensons qu'un tel produit pourrait subvenir aux besoins de nombreuses autres disciplines. Dans cette configuration en parcelles, l'information transmise par les attributs relatifs à l'utilisation du sol correspondrait avec les données de la statistique de superficie. Pourtant, il est presque impensable d'exécuter une agrégation sur les parcelles, mais nous pourrions nous servir de cette information pour distribuer spatialement les exploitations d'une façon plus réaliste. Il suffirait de prendre le centre de gravité de l'ensemble du domaine ou le centre d'un groupe de parcelles représentatif de l'exploitation, et d'appliquer ensuite les procédures d'interpolation et d'agrégation décrites au paragraphe précédent.

Peut-être se révélerait-il utile de délimiter des régions représentatives en se basant sur d'autres sources d'information. Les hypothèses conférant un grand rôle aux régions MS (de mobilité spatiale) se sont rapidement effondrées, ne trouvant aucun lien suffisamment significatif avec les données de l'agriculture (M. Gilgen, 1998). Nous pensons pourtant qu'il pourrait être intéressant de se renseigner auprès de centres collecteurs (moulin à céréales, laiterie, abattoirs…). De telles entreprises pourraient fournir, soit des données brutes (quantité de marchandises achetées, transformées, livrées…), soit des informations permettant de délimiter une sorte de bassin de producteurs. Il faudrait toutefois trouver un compromis sur les différentes sources explorées, de manière à ce que cette technique ne génère pas des régions trop grandes pour exprimer certaines particularités locales (Tessin…). Cette solution proposerait ainsi une taille intermédiaire, assez proche du niveau communal, mais suffisamment grande pour éviter des interventions ultérieures telles qu'une agrégation.

12.5 AUTRE TRANSFORMATION DES DONNÉES DU RECENSEMENT

La richesse des informations collectées par le recensement des entreprises du secteur primaire nous semble sous exploitée. La transformation de certaines observations sur une échelle monétaire ouvre de grandes possibilités d'analyses. Cependant, comme mentionné au paragraphe 2.4, ce nouveau type de données est dépendant des fluctuations de prix entrant dans le processus de transformation. Pour garantir la stabilité et la continuité des valeurs enregistrées, il serait intéressant de développer un système pour les convertir ou les normaliser. L'analyse en composantes principales s'est révélée infructueuse jusqu'à maintenant, mais il serait intéressant de persévérer dans cette voie.

13 CONCLUSION

L'objectif principal de ce travail consistait à élaborer une procédure d'agrégation de communes permettant la diffusion de données de la statistique agricole sous la contrainte de la confidentialité. Nous avons utilisé les fonctionnalités d'un système d'informations géographiques pour la sélection des communes selon leur attribut du nombre d'exploitations. De plus, nous avons pu déterminer les polygones adjacents aux germes et les comparer sur la base de leurs attributs et leur géométrie.

En outre, nous avons testé différentes méthodes, que l'on peut séparer en deux catégories : la corrélation sur les rangs et le calcul d'une distance euclidienne déterminent le voisin le plus ressemblant sur un unique facteur. Les systèmes utilisant une résolution multicritère appuient leur choix sur plusieurs indicateurs d'importances différentes, cette procédure permettant de trouver le meilleur compromis en fonction des diverses pressions (esthétisme, administration, fidélité à la réalité, etc.).

Dans un souci de généralisation, nous avons expérimenté une méthode hybride réunissant les avantages des deux systèmes rassemblés en tentant de minimiser les inconvénients. Ainsi, la combinaison multicritère + calcul de distance permet d'obtenir de bons résultats d'agrégation, tout en offrant des possibilités d'application à d'autres domaines de la statistique.

Enfin, nous avons pu estimer la marge de progression d'un tel prototype pour qu'il s'apparente à un outil intégral : nous avons pu déceler quelques limites de la méthode et proposer quelques améliorations et développements intéressants. D'autre part, nous avons pu évaluer dans quelle mesure une telle procédure correspond à une attente, et quelles sont les applications ou domaines pouvant s'appuyer sur cette expérience.

Nous nous sommes contentés d'employer directement les données fournies par l'OFS, mais nous pensons possible et certainement profitable l'utilisation de divers outils d'un SIRS en guise de prétraitement. En effet, la combinaison de quelque statistique agricole avec d'autres données à référence spatiale pourrait aboutir à la création d'une variable commune plus facile à cartographier ou de signification plus évidente.

14 BIBLIOGRAPHIE

Textes législatifs

- ***Loi fédérale sur la protection des données*** (LPD), Berne, juin 1992
- ***Loi sur la statistique fédérale*** (LSF), Berne, octobre 1992
- ***Ordonnance relative à la loi fédérale sur la protection des données*** (OLPD), Berne, 1993
- ***Ordonnance sur le relevé et le traitement de données agricoles ou ordonnance sur les données agricoles*** (ODA), Berne, décembre 1998

Polycopiés / Livres

- ***Méthodes d'analyse géographique quantitative***,
  HUBERT BEGUIN, Librairies Techniques, Paris, 1979
- ***ArcView GIS, The Geographic Information System for Everyone***, Manuel d'utilisateur, ENVIRONMENTAL SYSTEMS RESEARCH INCORPORATION (ESRI), USA, 1996
- ***Avenue, Customization and Application Development for ArcView***, Manuel d'utilisateur, ENVIRONMENTAL SYSTEMS RESEARCH INCORPORATION (ESRI), USA, 1996
- ***Structures des exploitations, Méthodologie des enquêtes communautaires***,
  EUROSTAT, Thème Agriculture, Sylviculture et Pêche, Ed. EUR-OP, Luxembourg, 1996
- ***Observatoire de l'agriculture durable***,
  Travail de diplôme, MARC GILGEN, Lausanne, février 1998
- ***Locational Analysis in Human Geography***,
  PETER HAGGETT, ANDREW D. CLIFF, ALLAN FREY, Second Edition, Edward Arnold Editions, Londres, 1977
- ***Algèbre linéaire***,
  Prof. THOMAS LIEBLING, Lausanne, 1996
- ***Probabilités et Statistique pour ingénieurs***,
  Prof. STEPHAN MORGENTHALER, Lausanne, octobre 1995
- ***Introduction à la statistique***,
  Prof. STEPHAN MORGENTHALER, PPUR, Lausanne, 1997
- ***Reflets de l'agriculture suisse, Edition 1998***,
  OFFICE FÉDÉRAL DE LA STATISTIQUE, Neuchâtel, 1999
- ***GEOSTAT, manuel de l'utilisateur***,
  OFFICE FÉDÉRAL DE LA STATISTIQUE, Neuchâtel, 2001
- ***Développement d'une méthode d'agrégation d'entités territoriales pour les besoins de diffusion et de protection des données de la statistique agricole***",
  Mémoire de diplôme de cycle postgrade, ROMAIN TORNAY, Lausanne, octobre 2001
- ***Modern applied statistics with S-PLUS***,
  W.N. VENABLES & B.D. RIPLEY, Second Edition, Springer Editions, New York, 1994
- ***Statistique, économie-gestion-sciences-médecine***,
  THOMAS H. & RONALD J. WONNACOTT, 4e édition, Editions Economica, Paris, 1991

Articles

Un tour d'horizon riche et séduisant
Communiqué de presse, OFFICE FÉDÉRAL DE LA STATISTIQUE, Neuchâtel, novembre 1997

Net recul de l'emploi dans l'agriculture entre 1990 et 1996
Communiqué de presse, OFFICE FÉDÉRAL DE LA STATISTIQUE, Neuchâtel, janvier 1998

Programme pluriannuel de la statistique fédérale pour les années 1999-2003
OFFICE FÉDÉRAL DE LA STATISTIQUE, Neuchâtel, 2000

Une nouvelle approche pour recenser les entreprises du secteur primaire
Communiqué de presse, OFFICE FÉDÉRAL DE LA STATISTIQUE, Neuchâtel, mai 2000

La statistique agricole n'a jamais été aussi importante
Communiqué de presse, OFFICE FÉDÉRAL DE LA STATISTIQUE, Neuchâtel, novembre 2000

Le secteur primaire toujours plus petit
Communiqué de presse, OFFICE FÉDÉRAL DE LA STATISTIQUE, Neuchâtel, septembre 2001

Qu'est-ce qu'un système d'information géographique ?
Manuel de présentation de GEOSTAT, OFFICE FÉDÉRAL DE LA STATISTIQUE, Neuchâtel, 1999

Solution agricole dans le domaine Online
Communiqué de presse USP, PSL, EMG, SAGmbH, Berne, septembre 2001

Etablissement des informations statistiques
Extrait du site web de l'OFS, OFFICE FÉDÉRAL DE LA STATISTIQUE, Neuchâtel, septembre 1998

Sites Internet

Agence d'Information Agricole Romande (AGIR), www.agirinfo.com

Agrigate AG, www.agrigate.ch/

AgriHebdo, www.agrihebdo.ch

EPFL, Institut de géomatique, Chaire de SIRS, http://dgrwww.epfl.ch/SIRS/index.fr.html

ESRI, www.esri.com ou http://esri-suisse.ch/ ou www.esrifrance.fr

Office fédéral de l'agriculture (OFAG), www.blw.admin.ch/f/

Office fédéral de l'environnement, des forêts et du paysage (OFEFP), www.umwelt-schweiz.ch

Office fédéral de la statistique (OFS), section agriculture et sylviculture, www.statistik.admin.ch/stat_ch/ber07/fber07.htm

Office fédéral de la statistique (OFS), section statistique de superficie, www.statistik.admin.ch/stat_ch/ber02/asch/fframe1.htm

Station fédérale de recherche en économie et technologie agricole de Tänikon (FAT), www.sar.admin.ch/fat/f/index.html

Comité "Oui aux accords bilatéraux", www.bilaterale.ch/f/

Bureau de l'intégration BFAE/BFE, www.europa.admin.ch/ba/expl/factsheets/f/index.htm

Union Suisse des Paysans (USP), www.bauernverband.ch/